

## Dangers of Euro-relevance

The European Commission has clarified aspects of its next Framework programme of research. An insistence on quick delivery of socio-economic benefits threatens the programme's success.

Here is relevance with a vengeance. Urged on by its member states and by the European parliament, the European Commission has at last produced a fifth Framework programme of research (FP5) that is so demanding of economic and social benefits that — reading between the lines — even some at the commission are unsure how, or even whether, it will deliver the goods.

As was clear from the long-delayed launch of the programme, in Essen last week, scientists seeking funding in FP5's four thematic programmes — life sciences, information technologies, industrial technologies, and energy and environment — will have to adopt what the commission calls a “problem-oriented approach”. It will no longer be enough, as it was in FP4, to propose a collaborative project in applied research. This time round scientists will have to present to the commission a problem of socio-economic significance, and show that they will approach it from as many angles as are necessary to solve it.

How this will work in practice takes some explaining, as the commission attempted to do in Essen. The “problem-solving” approach, explained one speaker, involves identifying a target problem (for example, getting wet surfaces to stick to each other) with advice from end-users (dental and medical communities), then bringing together the disciplines necessary to solve the problem scientifically (chemists, structural biologists, molecular biologists ...) while integrating the problems of production from the outset (include a bioreactor specialist). A successful consortium will probably include not only scientists but also representatives of other sectors such as social science or finance, from at least two EU countries (and more credit for more countries).

This level of organization will probably put off many scientists, especially those in academia, from trying to lead a project themselves. The commission defensively argues that FP4, which established the European networks that it now wants to exploit, gives European scientists the necessary head start to create the required contacts; but

it is unlikely that these contacts will suffice for the new ideas of FP5.

Some commission officials admit in private that they would not be surprised if they received a relatively low number of applications in response to the first calls for proposals, which will be issued mid-March with closing dates in early June. They expect that many scientists will prefer to watch the outcome of the first round before committing large amounts of time to preparing applications, especially when, as seems likely, they are not confident of the criteria for project eligibility. More openly, some commission officials admit that their own ideas of what specific programmes they might accept in the way of applications will only become completely clear once they see what sort of applications come in. They stated clearly in Essen that project evaluators and programme monitors will need training to operate the new approach.

EU member states demanded a radical change in the distribution of research money to improve Europe's poor performance in technology transfer. National governments can therefore take the credit if the new concept, which appears unprecedented, works. But they will deserve at least as much blame as the commission if, as seems likely, it does not.

Those concerned for the scientific future of Europe should give one cheer for the fact that a programme demonstrably successful at building networks and supporting fellowships in basic research, originally known as “Human capital and mobility” but now called “Increasing human research potential and the socio-economic knowledge base”, will continue and does not insist on practical problem-solving for support. But FP5's insistence on relevance combined with its managerial demands makes it significantly more remote from the realities of Europe's science base than its predecessors. That is a cause for significant concern. In the persistent absence of strong European alternatives, all the more reason, also, for European governments to keep their national support for basic research strong. □

## Where next on misconduct?

Tackling scientific fraud has generated its own problems, while others remain unaddressed.

Scientists are just a little bit more morally admirable than most groups of intelligent people. That, at least, is what C. P. Snow told a meeting of the American Association for the Advancement of Science 40 years ago. Whatever the truth of that assertion then, nobody aware of the record of fabrication, falsification and plagiarism in science could say the same now. What is more, some laboratories are so internally competitive that researchers are forced to keep their work close to their chest. Where that work is also difficult to replicate because of its experimental complexity or refinement, as is most particularly the case in biomedical research, you have the breeding ground for those grand misdemeanours that have led to hundreds of thousands of dollars of research grants being wasted, and, on some occasions, to even more costly legal battles.

Science is tough and is well able to survive such troubles. But public confidence in it risks being unduly undermined and the lives

of innocent individuals can be ruined. Following the lead of the United States and Scandinavia, the scientific world is confronting the situation — although the United States also provides warnings of procedures to avoid (see Briefing, pages 13–17). How to deal with accusations, how to look after the interests of the accused and of whistle-blowers, how to correct the scientific record: all of these are issues where examples have been set. High on the agenda now is to spread principles of good scientific practice.

But there is too much reliance on whistle-blowers: systematic spot checks on data and on publication records deserve more effort. Burgeoning costs in the United States require a major rethink and streamlining of the handling of misconduct cases. And there is also need for a systematic study, by surveys if necessary, of issues of credit misappropriation, exploitation and inappropriate pressures arising in laboratory hierarchies. In short, the community is making progress in fighting misconduct, but has far to go. □



# Anomalies in French blood inquiry over 'misleading' report

[PARIS] Documents from the mid-1980s reveal that the French agency that later carried out key investigations into the 'contaminated blood' affair was aware at the time that blood products were contaminated with HIV.

Staff at the General Inspectorate for Social Affairs (IGAS) not only failed to sound the alarm, but published a report in 1985 stating — misleadingly — that French blood products were of "satisfactory quality" and "much better than previously".

In an interview with *Nature*, Michel Lucas, who recently retired as head of IGAS, confirmed that this had been the situation at the time. But he denies that IGAS was at fault, and vigorously refutes allegations that the agency's involvement influenced his later handling of two major inquiries into the contaminated blood affair.

As head of IGAS, Lucas was the author of an inquiry carried out in 1991 into the decision-making process during the mid-1980s, when blood became contaminated with HIV. The results of the inquiry directly shaped both the current trial of three former ministers, and trials of more than 30 transfusion officials and government advisers.

A verdict is expected next week in the trial of former prime minister Laurence Fabius, former minister of social affairs and solidarity Georgina Dufoix, and her secretary of state, Edmond Hervé. They are charged with involuntary homicide and endangering life, and of negligence in handling the risk of transmission of AIDS in the blood supply in 1985.

The 1991 inquiry into the contaminated blood affair made no mention of IGAS's 1985 review of the National Blood Transfusion Centre (CNTS). This review endorsed the blood centre's commercial strategy — and made no mention of AIDS.

In cross-examination last week, Gérard Welzer, a legal adviser to Hervé, suggested that Lucas might have omitted mention of the 1985 report in his 1991 inquiry because he was "scared at some point to be questioned" about its apparently reassuring conclusions.

Lucas told a parliamentary inquiry in 1992 that the conclusions of the 1985 report concerning the quality of blood products were based on an analysis commissioned by IGAS from the National Health Laboratory. Had the IGAS investigator then had knowledge of contamination, he told the inquiry, "much greater vigilance would have been paid to the analyses carried out by the laboratory".

But documents obtained by *Nature* show that the rapporteur of the IGAS report, a Mme Broyelle, was in fact present at a meeting of the



**Lucas: headed investigative agency that failed to acknowledge its awareness of problems in 1985.**

Consultative Commission of Blood Transfusion on 20 June 1985, at which the contamination of blood products with HIV was discussed at length.

The minutes of the meeting state, for example, that "It must be known that the possibility of not having contaminated lots is very low ... under these conditions, it is indispensable that seronegative haemophiliacs are treated with heat-inactivated products."

According to the minutes, Broyelle, who has since died, "called the attention of the fractionating centres to new manufacturing procedures whose guarantees of safety vis-à-vis transmissible diseases need to be closely examined".

Asked to comment on the minutes, Lucas initially argued that the report was released in May 1985, before this meeting took place.

But the report remained open for comment by IGAS until the end of July 1985. The final report was sent to Dufoix on 19 September 1985 with a covering note from Lucas stating that the "development of the CNTS over the past few years has been extremely satisfying in technical terms".

Lucas subsequently suggested that the IGAS rapporteur might have failed to realize the importance of the discussions at the meeting. "Those who received information in the early 1980s didn't always realize its significance," he added, arguing that responsibility lies with those who gave the information and had a better knowledge of the consequences — and that the report's conclusions were not contested by CNTS officials.

Lucas says he had always instructed his inspectors "to bring to his attention any important facts that arose at meetings they attended". One reason the discussion at the 20 June 1985 meeting was not brought to his personal attention, he suggests, was that he

was in hospital during the summer of 1985.

Asked why IGAS's 1991 inquiry made no reference to the agency's 1985 report, Lucas said that the remit of the inquiry was merely to analyse the decision-making process in the mid-1980s. The IGAS report was therefore irrelevant, he argues, as it played no role in the decision making.

The report "would only have been relevant if IGAS itself had learnt that the blood was contaminated," says Lucas, arguing that, although Broyelle may have been aware of the dangers at the time, IGAS itself was not.

But the main reason, he adds, was that, if he had explained why the conclusions of the 1985 IGAS report were misleading, he would have been obliged to have accused the National Health Laboratory of having misled IGAS.

This would have conflicted with the remit of the 1991 inquiry, which was only to describe the course of events — not to seek to attribute responsibility to individuals.

Lucas says he was reassured — now, it appears, falsely — by the fact that the conclusions of the 1985 report were based on an analysis of the quality of blood products commissioned by IGAS from the National Health Laboratory. It was because of this, he adds, that he saw no reason to ask further questions.

It was only in 1991, he says, that he learnt that the analyses carried out by the national laboratory had not included testing the blood products for HIV, even though it had been informed that the products were contaminated with HIV.

Lucas argues that IGAS was misled by both the laboratory and the Directorate General of Health. "They should have told me [about the contamination] before the mission started."

Lucas also argues that any interpretation of the events must take into account the heavy workload under which IGAS was operating at the time. But he concedes that the conclusions of the 1985 report are ultimately "the responsibility of IGAS, and eventually mine".

Lucas also accepts that, when he became aware of the 1985 meeting during his 1991 inquiry, he realized that this could pose a potential problem for the agency. But he firmly rejects allegations that this influenced his actions in any way.

He dismisses allegations of conflict of interest as a bid to distract attention from the central allegations in the contaminated blood affair. Some protagonists are pursuing this aspect, he suggests, to support what he describes as an untenable argument, namely that in the affair "everyone is responsible and therefore no-one is guilty".

Declan Butler

# British drive for more women in science

[LONDON] Britain's science minister, Lord David Sainsbury, last week launched a project to promote women in science that has the long-term goal of seeing the same proportion of women in academic appointments as those recruited as undergraduates.

The varied strategy of the so-called Athena Project includes collecting and disseminating statistics, fostering personal and organizational development by raising the awareness of good practice in human resources, and facilitating and funding initiatives and awards.

Over the next four years, Athena aims to help achieve a 10 per cent increase in the number of women recruited to all levels of the academic community. In one project, higher education institutions will be invited to bid for grants for projects designed to improve the participation and promotion of women in science, engineering and technology, and to develop good practice.

Launching the project, Sainsbury said the British government was "well aware that we are not using all the available talent, and that the skill pool from which we draw our brightest academics could be enhanced by drawing on the wider pool of talented women".

A register of women in higher education, analysing their education, qualifications, positions and salaries, will form an important part of the project. Initially focusing on data for women in science, engineering and technology, the scheme will offer services



**Bullivant: seeking the support of industry.**

such as an information centre, resources and support for career development. The government believes that collecting and publishing gender-related statistics on issues such as different employment rates among institutions could provide a strong lever for change. It is hoped that the register will address issues such as training, career breaks, promotion and the profile of women in higher education.

Those responsible for launching the project acknowledge that the goodwill of the senior university managers, academics, politicians and industrialists who attended last week's meeting will need to be backed up by more money.

In addition to an initial £100,000 (US\$160,000) for the first year, newly appointed project director Susan Bullivant needs to find at least as much again.

Core funding currently comes from higher education and the government, but industry and learned societies are high on Athena's list of potential sources of additional funding.

Bullivant says she is looking for organizations that may share Athena's objectives. "We

might be able to help them in their strategic aims," she says.

Changing attitudes, cultures and practices within universities will not be easy. "It has to be done in a consultative and collaborative way," says Bullivant, who argues that the issue is about best practice in human resources, and so obtaining the best people irrespective of gender.

"An organization with a quality reputation is one where staff and students want to go," says Bullivant. "So the role of getting women is linked to major drives in the sector — to get the best people." Brian Fender, chief executive of the Higher Education Funding Council for England, agrees: "Our motive is that we want the best possible progression of education."

Athena builds on earlier work by the Scottish Higher Education Funding Council, which produced a series of guides to best practice in three key areas of women's involvement in academic science, engineering and technology: their access, participation and progress. The guides included details of successful projects and pilot studies in areas of concern.

One of the studies followed mentoring pairs at the University of Edinburgh, and concluded that mentoring increased assertiveness and networking skills in those being mentored, with both parties demonstrating increased confidence. One of Athena's aims will be to help establish mentoring schemes.

Natasha Loder

## Indian budget boost targets innovation, vaccines and genetics

[NEW DELHI] Funding for Indian science is to be increased next year to US\$2.44 billion — 3.6 per cent of the total national budget — under proposals submitted by the Bharatiya Janata Party government to parliament last week.

The government also announced three new schemes: a national fund for promoting innovation, a technology mission on vaccines, and the creation of a national bioresources board to conserve and manage India's genetic resources.

The science budget does not include \$399 million allocated for nuclear-power construction projects, including the development of a prototype fast-breeder reactor. "There is an increase in funding, though not substantial, and scientists are happy," says Valangiman Ramamurthi, secretary to the Department of Science and Technology.

As in the past, the three 'strategic' departments of defence, atomic energy and space get the lion's share (51 per cent) of the money (see Table 1). The Space Department,

which is developing a large cryogenic rocket engine, a second launch pad, and a high-power direct broadcasting satellite, gets an increase of 16 per cent. All increases are calculated over the final 1998–99 figures, rather than the significantly higher budget figures initially proposed.

The 32 per cent increase in the atomic-energy budget includes funds for advanced

plasma devices and a new national centre for applied mathematics and radio astrophysics, to be set up at the Tata Institute of Fundamental Research in Mumbai (Bombay).

The electronics sector is slated for a 33 per cent increase in research funding, reflecting government emphasis on information technology. Crop research will see its budget grow by 26 per cent.

One of the three new schemes is a national foundation, with an initial fund of \$5 million, to compile a national register of innovators and to help convert their work into business opportunities.

Finance minister Yashwant Sinha announced the launch of a technology mission to focus on "new vaccines that will revolutionize the medical and health systems".

The proposed national bioresources board, to be headed by the minister for science, will coordinate "policies, research, documentation and legal protection" of the country's genetic wealth.

K. S. Jayaraman

Table 1 Proposed Indian science budget for 1999–2000 (in US\$ million)

| Sector                  | 1998–99 (actual) | 1999–2000 |
|-------------------------|------------------|-----------|
| Defence research        | 550              | 663       |
| Space                   | 359              | 418       |
| Atomic energy           | 277              | 366       |
| Agriculture             | 229              | 288       |
| Industrial research     | 174              | 195       |
| Medical research        | 143              | 162       |
| Science and technology  | 132              | 156       |
| Non-conventional energy | 72               | 85        |
| Electronics             | 37               | 50        |
| Biotechnology           | 26               | 29        |
| Ocean development       | 26               | 26        |
| Total                   | 2,025            | 2,438     |



# US aid to Russian weapons labs defended

[WASHINGTON] The US Department of Energy (DoE) is fighting to protect its 'lab-to-lab' collaborations with weapons scientists in the former Soviet Union following a critical report from the General Accounting Office (GAO), an investigative arm of Congress.

The GAO report, published last month, argued that the Initiatives for Proliferation Program (IPP) is inefficient in getting money to scientists, and inadvertently supports work of military value to Russia (see *Nature* **397**, 643; 1999).

But Leonard Spector, director of the office of arms control and non-proliferation at the DoE, says the GAO's assertion that most of the \$62 million spent on the IPP since 1994 has gone to the department's US laboratories is wrong.

He also rejects the GAO's contention that the department is failing to screen the projects adequately to ensure they are of no military use to Russia.

Despite this, the DoE has said that it is addressing other criticisms in the report. These include a tendency to continue to support projects with no realistic prospects of commercial success, and the consumption of much of the money allocated to scientists by Russian taxes and institute overheads.

The GAO report was requested by Senator Jesse Helms (Republican, North Carolina), chairman of the Senate Foreign Relations Committee and a long-time opponent of aid to Russia. It also argues that none of the 400 IPP projects has so far achieved "long-term commercial success".

After the release of the report, which attracted widespread news coverage, the DoE may face an uphill struggle to win the \$30 million it has requested for the IPP in the 2000 financial year.

The programme began four years ago, evolving out of individual lab-to-lab activities initiated by scientists in the DoE nuclear weapons laboratories with their counterparts in the former Soviet Union.

Eighty-four per cent of its projects are associated with institutes in Russia, the remainder involving Belarus, Kazakhstan or the Ukraine. Its aim is to keep former nuclear, chemical and biological weapons scientists gainfully employed, while improving their long-term economic prospects by developing commercially useful products.

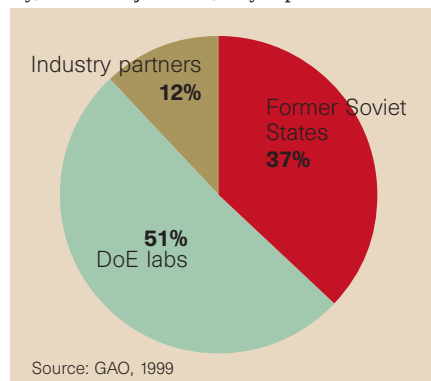
The United States pays for the projects, supporting a principal investigator at one of its own laboratories and a co-principal investigator and research team at the former Soviet institute. It also pays 'administrative costs' to a US industrial partner that will oversee the commercial exploitation of the projects.

The GAO's strongest criticism was that most of the programme funds — 51 per cent during 1994–98 — were disbursed to DoE

laboratories, with the target institutions receiving 37 per cent and the balance supporting the US industrial partners.

Spector says, however, that the GAO misattributed a large investment in telecommunications at the partner institutions to the DoE laboratories. "The [GAO] pie-chart is wrong, and the real split is about 50–50," he says. He added that the division was defensible, given that salary differences mean a US scientist costs about 20 times more to support than one in the former Soviet Union.

The DoE denies that it is helping Russia militarily — a charge the GAO backs up with examples mostly related to materials science. The report does say that laboratory officials told the GAO that "any improvement in materials" had potential military value. "We are very surprised with the charge that we are contributing to the Russian military capability, and we reject that," says Spector.



**Economic divide: DoE denies GAO's claim that most funds for joint work end up in its own labs.**

The report also expresses concern that scientists at some of the partner institutes have had links with countries the United States wants to isolate from Western military technology, such as Iran and Libya. But DoE officials point out that none of the institutes involved in the IPP were among those black-listed by the US State Department last year in response to nuclear-proliferation concerns.

The officials admit that the commercialization of IPP projects has been difficult, although they say that four years is too short a time to assess their full potential.

A handful of IPP projects have moved into early commercial exploitation. Pacific Northwest National Laboratory, for example, is working with the Khlopin Radium Institute in St Petersburg on a pharmaceutical product that is already on sale in the city.

"It's not impossible" to make a commercial success of the projects, says Patricia Godoy-Kain, IPP programme manager at Pacific Northwest, which has recently embarked on a large number of IPP projects, mainly with partner institutes previously engaged in biological-warfare research.

The Congressional response to the report has been predictably mixed, with Democrats supporting the programme and some Republicans doubting its value. Helms said the programme's future would be threatened unless the findings were addressed.

Spector is optimistic. "I believe that the administration and the Congress are both very much behind this kind of interaction." He adds, however, "that there is room for improvement".

Colin Macilwain

## Japanese institutes face greater autonomy

[TOKYO] Prominent basic research institutes in Japan, such as the Institute of Space and Astronomical Sciences, appear likely to be added to the list of organizations due to be transformed into semi-autonomous 'agencies', with greater managerial independence.

The Ministry of Education, Science, Sports and Culture (Monbusho) is considering whether the National Research Institutes for Joint University Use should be included in reform plans aimed at improving the nation's administration (see *Nature* **389**, 897; 1997). These institutes provide facilities and equipment for domestic and international researchers.

The institutes — which include the High Energy Accelerator Research Organization and the National Institute of Genetics — would become semi-autonomous, alongside research institutes attached to science-related ministries, such as the National Institute for Environment Studies.

In proposing the move last month, the government said that the focus of these bodies on research showed that their role in university education was minimal, and that they were therefore equivalent to the national research institutes.

The plan, which is likely to be formally adopted next month, will give each institute its own management system and an external assessment body to review its performance. The new status will give the institutes greater flexibility in funding, for example by allowing them to carry over unspent research funds to the next fiscal year.

Concern had been raised over the plan's emphasis on financial performance, but these targets have been revised and would no longer apply to long-term basic research. Although the restructuring will include an accounting system comparable to that of private companies, it will not require cost-related performance from research institutes.

Asako Saegusa

# Collapse of talks on safety of GMO trade

[LONDON] Nine days of talks on a global protocol to regulate international trade in genetically modified organisms (GMOs) broke down last week, even though at one point agreement appeared within sight. Negotiations are not expected to resume before the next conference of the parties to the United Nations biodiversity convention in May 2000 at Nairobi.

The main reason for failure appears to have been the insistence by grain-exporting countries — led by the United States and known as the Miami group — that countries be allowed to export genetically modified commodities such as food, pharmaceuticals and animal feed without seeking permission from an importing country.

The Miami group was also keen that the protocol should not conflict with existing international trade agreements. In contrast, European Union (EU) member states, most developing countries present, and environmental groups, wanted the protocol to be an independent legal instrument.

Developing countries also wanted the socioeconomic impacts of GMOs to be taken into account during any assessment of their environmental risks, as well as provision in the protocol for compensation in the event of accidents involving the transport of GMOs.

Earlier during the talks, agreement had appeared close when the chair, Veit Koester, director of the National Forests Agency of Denmark, produced a compromise text for a protocol satisfying these demands.

The compromise text also satisfied the Miami group's demands that only genetically modified organisms — and not commodities — would need permission before being shipped to an importing country.

Developing countries, the EU states and



the Miami group appeared poised to agree on a later version of this text, which included the possibility of some genetically modified commodities being subject to import permission. But the talks collapsed when the Miami group refused to budge on the article specifying that the protocol should not conflict with existing international trade agreements.

Some delegates have suggested that progress in the talks may have been hindered by the predominance of scientists among the delegates from smaller countries, for example from Africa, with little experience in international negotiations.

But Sateaved Seebaluck of the Ministry of Local Government of Mauritius says that, although smaller countries were unable to field the full range of expertise, the emergence of the African group as a single block has changed this. "Now we have access to as many scientists, lawyers and diplomats as we need."

The lack of agreement has intensified an increasingly significant conflict between

trade and environmental concerns. A future agreement will be a key test of the UN biodiversity convention, which is seen by its supporters as the only international agreement in which environmental protection is given the same importance as trade.

According to Kimo Goree, managing editor of the *Earth Negotiations Bulletin*, "the protocol could prove to be the convention's saving grace, or a nail in its coffin as resolution of such trade issues will be key to its final role". The bulletin was the only publication allowed access to closed negotiation sessions.

But the lack of agreement still seems to have satisfied both sides, at least for the time being. Representatives of industry left the meeting in Cartagena, Colombia, happier than when they arrived. They have always favoured international guidelines on trade in GMOs, as opposed to a legally binding protocol, which they consider would be harmful to trade.

The lack of agreement also represents the first major success of industry's more conciliatory approach to international environmental negotiations, which in the past has been much more confrontational.

Unlike at the climate change negotiations that resulted in the Kyoto protocol in 1997, industry played an active role in biosafety preparatory meetings, and had a much closer relationship with key governments, including the United States.

Despite losing the support of Argentina, Chile and Uruguay to the Miami group, but gaining the support of South Africa, developing countries and environmental groups, on the whole, are not displeased with the outcome. Many of them felt that they were better off not signing what they privately considered to be a weak agreement.

Ehsan Masood

## NASA plans two-stage Hubble repair to keep the data flowing

[WASHINGTON] The US space agency NASA may repair the Hubble Space Telescope earlier than scheduled, to prevent a failure in its guidance system that would temporarily halt scientific observations.

If the plan goes ahead, an astronaut crew will replace the telescope's gyroscopes in October, rather than waiting until the planned 'servicing mission' in June next year.

The same crew would then return a year later to install a new scientific camera and to perform other repairs scheduled for the June 2000 mission. An emergency rescue is needed following the partial failure in January of one of the telescope's six gyroscopes, which keep it orientated in space. Hubble needs three gyroscopes to function normally, with the rest as backups.

But, with two having already failed

completely, an electrical malfunction caused a third gyro to behave erratically, and it was shut down by engineers at NASA's Goddard Space Flight Center in Maryland.

According to David Leckrone, the Hubble project scientist at Goddard, there is no way to predict when a similar failure might occur in another gyro. A fourth loss would automatically put Hubble into a 'safe' mode, where it would remain protected but stop observations.

Hubble managers endorse the October mission, but NASA's space shuttle and science offices have yet to agree to the plan.

Delays with the Russian 'service module' have left shuttle managers juggling their launch schedule. The module was to have been delivered to the space station in September, but now may not fly until

November or later. These Russian schedule slips left the shuttle with no mission this autumn, allowing the Hubble rescue.

NASA science managers have a different concern. Splitting the repair mission would give twice the opportunity to damage or contaminate the telescope.

But, says Leckrone, the planned June 2000 repair, which will install an Advance Camera for Surveys and cryogenic cooling system, and update computers, "was shaping up to be quite complex," with six space walks. Spreading the work over two missions makes it more manageable, he says.

NASA administrator Daniel Goldin told a congressional committee last week that NASA wants to decide soon. Leckrone says it may be a "couple of weeks" before the agency can commit to the plan.

Tony Reichhardt

# Germany aims for competitive equality

[BONN] Biotechnology researchers, east German scientists, women and young scientists are to be the main beneficiaries of growth in the German research budget, which the Social Democrat–Green government has promised to double in five years.

“But these groups will have to compete strongly against each other for the new funds, and will in many cases have to work in collaboration with industry,” Wolf-Michael Catenhusen, Germany’s junior minister for research, told *Nature* last week.

There may also be less money for space science, as funds are diverted to pay for Germany’s participation in the International Space Station (ISS) and the development of the European launcher Ariane 5.

Catenhusen served for eight years as chairman of the parliamentary science committee in the late 1980s and early 1990s.

Revealing details of the new government’s plans, Catenhusen said that Germany’s highly successful BioRegio competition, in which regions compete for special access to part of the existing biotechnology budget, will be continued (see *Nature* **379**, 759; 1996). The three winners will have access to an additional DM150 million (US\$85 million) from 1999 to 2003.

The plant genome programme GABI, launched last year by the previous research minister Jürgen Rüttgers with an undefined budget, will get DM5 million this year and between DM20 million and DM25 million next year. In the future, this may rise to DM30 million. Collaboration between academics and industry will be given priority.

Direct subsidies for east Germany, whose science base was rebuilt after reunification, will end. Some east German researchers were temporarily funded through the former *Wissenschaftlicher Integrationsprogramm* fund. But a new competition, based on BioRegio and known as InnoRegio, will promote innovation in the new *Länder* (states).

East German researchers have been so well supported in recent years that they are now considered good enough to compete with west German scientists for conventional funds, says Catenhusen. “Any deficits between east and west are [now] due to a lack of industrial research [in the east],” he says.

InnoRegio will begin next year with funds of DM30 million, expected to rise to DM70 million the following year. It will reward the consortia of academic scientists and industry — particularly small and medium-sized enterprises — that present the best plans to develop new technologies. The scheme’s planning phase will be supported with DM5 million put aside from this year’s research budget.

The special programme for universities, called the *Hochschulsonderprogramm*, which



**Catenhusen: boosting industrial research.**

the federal government set up in the late 1980s to address problems caused by mushrooming student numbers, will not be extended beyond 2000. But the government will take over two parts of the programme that fit its own priorities: special support for women in science and for the promotion of young scientists.

The government has already put aside DM7.5 million this year for a programme called ‘Equality of opportunities for women in education and research’, and intends to increase this annual sum to DM12 million in 2000 and 2001. A special programme for young scientists is being planned, but details are not yet available.

Government-level discussions about making university hiring policy more flexible are also considering whether to revise rules that bar non-tenured scientists from working at a university or research institute for more than five consecutive years (see *Nature* **396**, 396; 1998).

“The structure of our conditions of employment no longer matches the way research is conducted in practice,” says Catenhusen, “in particular, [by] young people who have research grants which do not fit neatly into blocks of five years.”

He does not want researchers to become permanent fixtures in a particular institute, he says, but institutes should be able to extend employment beyond five years where a portion of grant money remains to be used.

Germany’s 16 national research centres, which carry out large-scale research, have a budget top-up of DM105 million, to be distributed competitively between them-

selves and any partners with whom they forgo links.

About half of this will go to the ‘strategy fund’ created two years ago by the centres’ umbrella organization, the Helmholtz Society, to fund projects in areas that society identifies as priorities. The rest will fund programmes supporting innovation, the development of new research areas and ‘networking’.

Catenhusen believes the DM15 million ‘networking fund’, for projects involving partners from industry or another research institute, such as a Max Planck institute, is particularly important. “National research centres have a tendency to operate in isolation,” says Catenhusen, “and this extra money should encourage them to look beyond their own walls.”

One loser in the distribution of the new bounty will be space science, whose budget is being held down and may be threatened more substantially in the next few years.

Research minister Edelgard Bulmahn told parliament last week that she intends to honour all international commitments to the ISS and Ariane 5. But she argued that the government should give her extra money to allow her to pay all the bills.

Without any extra money, the DM300 million domestic space science programme may be the only fund not yet fully committed, and could therefore be used to bail out the international programmes, according to ministry sources.

The domestic space budget has already seen a 10 per cent cut since 1994, with funds for space science and astronomy falling from DM120 million to DM80 million.

Reinhard Genzel, a director at the Max Planck Institute for Extraterrestrial Physics in Garching, says: “If money is really so tight, then the government should look very carefully at the ISS commitments and consider whether some of them could indeed be revised.” Space science “will suffer terribly if cuts are deepened,” he says. Alison Abbott

## Head of German particle physics lab dies

[MUNICH] Björn Wiik, director of Germany’s national research centre for particle physics (DESY), died last week after an accident at home.

The 62-year-old Norwegian physicist joined DESY in 1972. With three colleagues, he provided the first experimental proof of the existence of the elementary particle gluon using DESY’s Positron–Electron Accelerator.

His death comes at a critical time for the future of TESLA, a 33-km superconducting linear accelerator that Wiik hoped to build

at the DESY site in Hamburg by 2010. Plans for TESLA are being developed by 38 institutes in nine countries — including China, Russia and the United States — but financial support has not yet been assured.

Wiik also led the planning and construction of the Hadron–Electron Ring Accelerator, Germany’s biggest research facility, which went into operation in 1992. Wiik became head of the DESY in 1993.

The research centre’s directors have said that TESLA should be continued “as Wiik would have wished”.

A. A.



# South Africa stands firm on AIDS drug . . .

[CAPE TOWN] South Africa's minister of health, Nkosazana Zuma, has refused to alter her decision not to pay for a pilot programme administering the antiviral agent AZT to HIV-positive pregnant mothers. This is despite increased threats of a boycott of the World AIDS Congress in Durban next year.

New findings released in Chicago last month by UNAIDS, the joint United Nations programme on HIV/AIDS, and based on trials with 1,357 mainly breastfeeding women in Uganda, South Africa and Tanzania, indicate that a one-week treatment of mother and child with both AZT and 3TC (lamivudine), beginning at the start of labour, could reduce transmission by 37 per cent.

The president of the International AIDS Society, Mark Wainberg, visited South Africa at the end of January in what seems to have been a futile attempt to persuade Zuma to reverse her earlier decision.

Although he left the country "confident and encouraged" that the new information would cause a change of mind, Zuma has now confirmed that the government will not change its mind on a pilot study for AZT treatment, which it refused to fund last October (see *Nature* 396, 504; 1998).

The ministry has argued that the pilot programme would create unmettable expectations if funds were not available to follow it up with full implementation. But the annual base cost of implementing a short-regime programme nationally would be around R14 million (US\$2.3 million), from a R20 billion



**At risk: but one week's AZT treatment would cut maternal transmission of HIV by 37 per cent.**

annual health budget.

Additional costs arise from counselling and the provision of formula milk for mothers to feed their babies in lieu of breastfeeding. Glaxo-Wellcome is offering AZT to the South African government at 75 per cent below the price for developed countries, and has offered to hold that price for five years.

Along with replacing a four-week regime (which reduces transmission by 50 per cent) with a one-week one, this has reduced the initial cost estimates by a factor of five.

"We are disappointed. We cannot understand why Zuma is ignoring the data," says Glenda Gray, director of the Perinatal HIV Research Unit at Chris Hani Baragwanath Hospital in Johannesburg, one of the sites of the UN trials.

The ministry has continued to argue that it cannot find the funds for AZT treatment, despite counter-arguments that it is more cost-effective than treating HIV-positive

babies. Babies born with HIV account for about 20 per cent of new infections in South Africa each year, and the daily cost of paediatric care in a state hospital is R400.

Wainberg also called on South Africa's president, Nelson Mandela, its deputy president, Thabo Mbeki, and its minister of finance, Trevor Manuel, to recognize that perinatal treatment is the most cost-effective approach, and to make funds available for this. "South Africa must consider itself to be at war," he said. "Its number one enemy... is not some neighbouring country threatening its borders. It is HIV."

This is a thinly veiled reference to the country's decision last year to spend R30 billion on upgrading its defence force. But Vincent Hlongwane, Zuma's press secretary, replies that Zuma's position is fully supported by the cabinet, which has apparently made a policy decision to channel all disposable funds into preventative measures.

The International AIDS Society met with South African and international AIDS researchers in Chicago last month to discuss growing calls for a boycott of the 13th World Aids Conference, scheduled to be held in Durban next year.

Wainberg said they had agreed to oppose a boycott at this stage, but warned that if there were no signs of a change of policy by April (when he is to visit South Africa again) "it may be time to consider another course of action".

Local AIDS activists have not yet come out in support of a boycott because the conference is still 18 months away. But AIDS Consortium representative Mark Heywood has indicated that they will reconsider their position later in the year.

Meanwhile, the Centre for Applied Legal Studies and the National Association for People with AIDS are planning legal action to challenge Zuma's decision.

The Western Cape provincial government, which is controlled by the National Party, went ahead with the pilot programme in defiance of Zuma's policy. The programme was launched in January in the township of Khayelitsha, Cape Town, with the reported support of the African National Congress's leader in the province, Ebrahim Rasool.

But Zuma's handling of the issue does not seem to have affected her popularity within her party at national level. She is third on their national list for this year's general election, outranked only by party president Thabo Mbeki and deputy president Jacob Zuma, her former husband.

There is even speculation that she might be appointed deputy president in Mbeki's new government, ahead of both Jacob Zuma and the leader of the Inkatha Freedom Party, Mangosuthu Buthelezi.

Michael Cherry

## . . . while forecasting benefits from NASA funds

[CAPE TOWN] The US space agency NASA has agreed to locate a satellite laser ranging system (SLRS) in South Africa. The facility, the first on the African continent, will be sited at the Radio-Astronomy Observatory at Hartebeeshoek in the North-West province.

The project is one of several cooperative scientific ventures jointly agreed by South Africa and the United States after a meeting last month of the US-South African Binational Commission in Cape Town, led by US vice-president Al Gore and South African deputy president Thabo Mbeki.

The commission's science and technology committee met under the

joint chairmanship of Ben Ngubane, South Africa's newly reappointed minister of arts, culture, science and technology, and Neal Lane, science adviser to the US president, Bill Clinton.

South Africa will bear the operating costs of the SLRS, which has a value of \$10 million. The system, which allows for accurate assessment of satellite altitude, will be used for geodetic research, as well as increasing the capacity to make more reliable long-range climate forecasts.

The US National Oceanic and Atmospheric Administration and the South African Weather Bureau plan to collaborate on these forecasts. The last El Niño event did not materialize

in South Africa, as it was offset by a warm upwelling in the Indian Ocean that could have been predicted by an SLRS.

"The new system will fill an important data gap for the African continent," says Rob Adam, South African deputy director-general for science and technology, adding that he is "very positive" about the science and technology panel's deliberations.

Other ventures already in progress on which agreements were finalized, many of which concern science education, include the launching by NASA last week of a microsatellite, built at the University of Stellenbosch in a project that included the training of 50 postgraduate students. M.C.



# Canadians protest over rare species bill

[MONTREAL] More than 600 Canadian scientists have written to the prime minister, Jean Chrétien, demanding that planned legislation to protect endangered species is based on scientific criteria. The signatories include Nobel laureate Michael Smith and 13 fellows of the Royal Society of Canada.

At present, Canada has no such federal legislation, and its previous draft bill (Bill C-65), which died in 1997 because of an impending general election, was considered seriously flawed by the scientists.

The letter complains that the government has ditched two vital factors from its proposed new legislation: habitat protection, and the exclusive right of scientists to identify and list species at risk.

By giving the cabinet the power to override scientific decisions on species that should be on the list, the government would let political considerations interfere with the listing process, the letter adds.

"This is unacceptable," say the scientists. "Scientific findings — and scientific findings alone — should determine if a species needs to be listed as endangered. Scientists, by their expertise and independence, are better situated than politicians to judge the state of a species."

They add: "Canada's endangered species are too imperiled, too close to extinction, and too precious to be held hostage to lobbyists, political manipulation, or simple ignorance."

Jamie Smith and Geoff Scudder of the University of British Columbia's Centre for Biodiversity Research and David Schindler of the University of Alberta's biological sciences department were primarily responsible for drafting the letter, which was posted on the

web site of the UBC biodiversity centre (<http://www.zoology.ubc.ca/biodiversity>).

The scientists' concern has been heightened by the government's removal of the voting rights of most of the non-governmental scientists on the supposedly independent Committee on the Status of Endangered Wildlife in Canada (COSEWIC) charged with identifying and listing such species.

Bill C-65 would have given the federal cabinet the power to override the committee's list. "This would reduce COSEWIC to merely recommending that a species be listed, and would mean that species which are scientifically known to be at risk could be legally neglected," say the scientists.

They point out that two previous letters from scientists, in 1995 and 1997, emphasized that species at risk cannot be protected without protecting their habitats.

"These habitats can be geographically dispersed and are not confined within political boundaries, but must each be effectively protected to ensure a species' well-being," their letter argues. "Over 80 per cent of the species listed by COSEWIC are at risk because their habitats are threatened, and more species are being listed yearly (from 291 to 307 this year alone)."

The scientists argue that the new bill must improve on Bill C-65, which would have protected the habitats of fewer than half of Canada's endangered species. It also failed to protect the habitats of most species that cross Canada's international borders.

The letter describes this omission as being "radically at odds with Canada's treaty obligations" and a guarantee that Canada would contribute to the decline of species



**Big bird:** protestors want scientists to decide if species like this whooping crane are endangered.

elsewhere. "To be effective, the new bill must create a national legal standard for habitat protection for all species, not fragments of protection for a few species."

Amir Ataran, a biologist and lawyer who has coordinated the group, says that powerful lobbyists for industries such as forestry and mining oppose any legislation that threatens their interests.

Ataran says another factor is that in Canada the protection of endangered species is an area of overlapping jurisdictions between provincial and federal governments, and this can lead to conflicting interests. While the federal government does not want to offend the provinces, its priorities are social benefits and national unity, so legislation for endangered species becomes a "sacrificial lamb".

According to Ataran, the scientists hope that, by going public, they will put pressure on the federal government and force it to modify the new bill.

David Spurgeon

## Varmus wish list guides senators in fight for biomedical funding

[WASHINGTON] A key US senator says he will push for considerably more money for the National Institutes of Health (NIH) in 2000 than the 2.1 per cent increase President Bill Clinton has requested of Congress.

Senator Arlen Specter (Republican, Pennsylvania), chairman of the Senate appropriations subcommittee responsible for approving NIH funding, said last week that he would use as a "guidepost" in drafting next year's spending bill a "professional judgement" budget provided by NIH director Harold Varmus.

Specter said Varmus had told him last month that the NIH could usefully spend an increase of 24 per cent, raising its budget by \$3.7 billion to \$19.3 billion.

But, when asked by Specter about the figure at a hearing of the subcommittee last week, Varmus emphasized that it was based on a complete lack of constraints — "what

we could do under optimal conditions". It was therefore considerably higher than the sum the NIH actually requested of Clinton's Office of Management and Budget last autumn. That figure was a 10 per cent increase, according to an NIH official.

In the budget he submitted to Congress on 1 February, Clinton requested considerably less than NIH had suggested, asking for an increase of \$320 million, or 2.1 per cent (see *Nature* 397, 377; 1999).

It would be extraordinarily difficult for Congress to land a \$3.7 billion increase for NIH, as it is restricted by caps on discretionary spending in a 1997 budget law. These caps would force Congress to raid other agencies in the large spending bill of which NIH is a part in order to provide NIH with any substantial increase.

But, as budget surpluses begin to mount up, the political will to keep the budget caps

in place seems to be eroding. Both parties in Congress are looking for ways to undo the caps, possibly creating the opportunity for yet another generous NIH increase.

Apparently pointing in this direction, Specter introduced a resolution in January declaring the Senate's sense that biomedical research spending should be boosted by \$2 billion in the coming year.

And John Porter, chair of the House subcommittee that funds NIH, echoed Specter's interest in Varmus's 'professional judgement' budget in an interview with *Nature* last week.

Porter said he intends to keep NIH on course to double its budget over five years. That would require a 15 per cent increase in 2000, like the 15 per cent hike Congress gave the agency in 1999.

Both Specter and Porter are urging Congress to lift the caps.

Meredith Wadman

## \$8bn Everglades restoration plan gets independent review

[WASHINGTON] The US Secretary of the Interior, Bruce Babbitt, agreed last week to set up an independent standing advisory panel to review the science behind an \$8-billion programme intended to restore the Everglades ecosystem in south Florida.

The interior department and the Army Corps of Engineers had already been considering such a panel, but a recent letter by Stuart Pimm of the University of Tennessee and several other prominent ecologists called attention to problems with the restoration plan, and pointed to the need for additional outside review (see *Nature* 397, 462; 1999).

The panel, to be appointed by Babbitt, will include scientists with a range of expertise from ecology to hydrology who are not involved in Everglades research. Babbitt will seek recommendations on membership from the National Academy of Sciences and federal and state agencies.

## Satellite sets out to track down space junk

[BOSTON] An instrument launched last week on the Advanced Research and Global

**Observation Satellite is to circle the Earth for three years, studying orbital debris that can pose a hazard to spacecraft. SPADUS, as the space dust instrument is called, is designed to measure the mass, speed and trajectory of particles too small to be tracked from the ground.**

Scientists will be able to differentiate between man-made 'space junk' and cosmic dust produced by comets, meteoroid impacts and other natural processes. "This is the first active experiment where you can separate these two phenomena," says John Simpson of the University of Chicago, where SPADUS was designed.

## University pay dispute could disrupt exams

[LONDON] University staff in Britain are threatening to take industrial action if they do not receive an "acceptable response" by 1 April to demands for a significant salary increase. There could be a one-day strike, the disruption of examinations and student admissions, and a boycott of "all bureaucratic procedures".

The Association of University Teachers, the union that represents staff in their pay negotiations with the universities, claims that salaries for academic and related staff in the old universities have fallen behind other

professional groups by more than 35 per cent in recent years. They are seeking a 10 per cent pay increase as "the first step in catching up".

## Democrat warns NIH on stem-cell research ethics

[WASHINGTON] The director of the US National Institutes of Health (NIH), Harold Varmus, was warned last week by a liberal Congressman that he should not underestimate the "good faith" of the opponents of stem-cell research. The warning came from congressman David Obey (Wisconsin), the senior Democrat on the House Appropriations Committee.

Obey said the NIH must "lay out in considerable detail and with considerable force" its plan to ensure that the ethical issues in stem-cell research "are not stuck in the caboose on the train". About 80 members of Congress have written to Donna Shalala, the Secretary of Health and Human Services, protesting about her department's decision to fund the research (see *Nature* 397, 639; 1999).

## US panel backs electron laser to probe matter

[WASHINGTON] The United States is about to take the first step on a long road to the construction of a high-powered 'fourth

generation' X-ray light source. It would use a linear accelerator-driven free electron laser to probe matter in far greater detail than existing synchrotron light sources.

The Department of Energy's Basic Energy Sciences Advisory Committee last week accepted a broadly positive report from a sub-panel which called for preparations for the building of such a machine, starting with research programmes into the necessary accelerators and lasers.

These research programmes would lead to the construction of an electron laser test facility at the Stanford Linear Accelerator Center in California and, eventually, to an advanced X-ray source providing 8–20 keV X-rays to scientific users. But the panel cautioned that a "compelling and rigorous" scientific case for such a facility is yet to be fully developed by the scientific community.

## Argentina and Chile to collaborate in Antarctic

**[LONDON] The presidents of Argentina and Chile have issued a statement expressing their support for "furthering joint scientific undertakings in the Antarctic" and establishing protected areas "by common agreement".**

**In a declaration issued to commemorate**

**the fortieth anniversary of the signing of the Antarctic Treaty, Carlos Menem of Argentina and Eduardo Frei of Chile described the treaty as "a valuable international effort to promote peace and science". They said they were keen to enhance programmes currently under way in scientific research and associated logistical aspects, "thus enabling the optimization of resources earmarked for activities in the Antarctic".**

**The two presidents stated that they shared a desire to protect the Antarctic and its ecosystem, and were both keen to promote the treaty as a necessary means of preserving the continent "for the benefit of generations to come".**

## US acts against 'nuclear cooperation' with Iran

**[MOSCOW] The United States has imposed sanctions on 10 Russian organizations it suspects of defence and nuclear cooperation with Iran. The sanctions will prevent scientific exchanges with the United States and will halt supplies of equipment.**

The allegations have been vigorously denied. Viktor Chernomyrdin, Russia's former prime minister, described them as "far fetched" and said that the United States had failed to provide documentary evidence of similar allegations in the past. Pavel

Sarkisov, rector of the Mendeleyev Chemical Technology Institute in Moscow, one of the organizations affected, said his institute has no agreements with Iran.

## Seaborg, discoverer of plutonium, dies

**[LONDON] Glenn Seaborg, the Nobel-laureate chemist who discovered ten elements — including plutonium and one that bears his name — has died at his home near Berkeley, California, aged 86. He had suffered a stroke last August.**

**Seaborg, a member of the Manhattan Project and former chancellor of the University of California at Berkeley, was chairman of the US Atomic Energy**



**Commission, the predecessor of the Department of Energy, between 1961 and 1971.**

**Throughout his research career, Seaborg also championed science education, nuclear arms control, and international cooperation in science.**

# Science comes to terms with the lessons of fraud

Although the incidence of proven scientific fraud remains low, several high-profile cases have convinced the research community of the need for effective action, in particular by enforcing codes of good laboratory practice.

Until a few decades ago, fraud in science was generally believed to be uncommon and, when it happened, it was perceived as the action of the deranged. Science is supposedly self-correcting, and so the discovery of fraud was seen as inevitable. However, as laboratory life has become more competitive, and especially where experiments are difficult to replicate, fraud and other types of serious misconduct have become less rare.

Often, according to those who have studied the phenomenon, the motivation has not been to falsify but to cut corners in reaching a conclusion that the fraudster genuinely believes to be true. In the trusting world of science, such behaviour is hard to prevent.

But the growing problem of scientific misconduct brings with it increased disruption, more money and effort spent clearing up the mess, and public distrust of science. So it has become an increasing priority for research agencies worldwide to find effective ways of rooting out misconduct and, where appropriate, dealing with it firmly. They are also seeking effective preventive measures — for example, by encouraging adherence to clear principles of good scientific practice.

US research organizations have many years' experience of attempting to deal with the problem, which seems to be particularly prevalent in biomedicine. Their European counterparts have been slower off the mark, but are now starting to address the issue in earnest. In some cases this is the result of domestic experience; in others, an extension of demands for responsibility and accountability in all areas of public expenditure.

As well as exploiting the legwork that has been done in the United States — for example, by using definitions of exactly what constitutes misconduct (see box, page 16) — European research institutions are trying to avoid the mistakes of the US systems, for example in allowing investigations of allegations to drag on for years. They are also con-



centrating increasingly on finding ways to avoid fraud, for example by trying to remove both the motivation and the opportunity.

In the process, the concept of self-regulation by the scientific community remains sacred on both sides of the Atlantic. "But the right to self-regulation is not sacred," says Bruno Zimmermann, head of research affairs at the Deutsche Forschungsgemeinschaft, Germany's main grants agency. "It must be earned by showing the public that we are ensuring that research is being carried out in full honesty."

## How big is the problem?

The extent of scientific misconduct of the 'fabrication-falsification-plagiarism' type is hotly — and usually rather unscientifically — debated in both the United States and Europe. Some scientists fear that publicized cases are merely the tip of an iceberg. Others remain convinced that the overall incidence remains low, and even those with direct experience of misconduct cases are often optimistic.

Eberhardt Hildt is the young postdoctoral fellow who blew the whistle on what appears to be Europe's biggest fraud case, where two German cancer researchers published scores of papers that include apparently fabricated data. He says he still has faith in the fundamental honesty of science. "Otherwise I would not have decided to stay in science." Hildt now leads a research group at the University of Munich.

But Imogen Evans, research strategy manager of the UK Medical Research Council, points out that "the evidence on misconduct is impossible to interpret". She suggests that "the question of how widespread scientific fraud might be is unanswerable".

It is even difficult to gather accurate data on the number of research misconduct cases that have come to light, as funding agencies and research organizations have different procedures for recording cases. At some US government funding agencies, for example, the only way to learn about misconduct cases hidden in archives is to request documents under the Freedom of Information Act — a time-consuming process that can produce an incomplete picture, and require a court decision for the release of critical details.

The Office of Research Integrity (ORI), which addresses misconduct in projects funded by the US National Institutes of Health (NIH), has the most comprehensive methodology for reporting cases. It publishes regular newsletters that report the final decisions on cases, including the name of the wayward researcher, his or her institution, and the punishment meted out. This newsletter can also serve as the lone corrective device for published fabrications, as fears of lawsuits mean that some journals balk at retracting faulty articles (see box, page 15).

The ORI received about 1,000 allegations of misconduct between 1993 and 1997. It closed 150 investigations, deciding that misconduct had taken place in almost exactly half. Of the 76 researchers found to have committed misconduct, 71 per cent were barred from receiving federal research funds for between 18 months and eight years. Eighty per cent of cases involved the falsification of research data. The total number of NIH grants issued during the same five-year period was more than 150,000.

In Europe, only Scandinavian countries have analogous national committees that can collate such statistics, and all come up with relatively low numbers. The Danish Committee on Scientific Dishonesty was the first to be established, in 1992. Its responsibility was initially restricted to biomedicine, like the ORI, but its remit was extended to cover all sciences at the beginning of this year. The committee has considered only 25 cases in its seven years of existence, and found substance in only four of them.

The number of reported cases of misconduct is therefore relatively small compared to the total volume of scientific publications each year. The Science Citation Index listed nearly one million articles last year published in 5,600 selected journals — and this is probably only one tenth of the total number of scientific journals published worldwide.

But the few surveys that have been carried out suggest that the level of misconduct may

**This Briefing was written by Alison Abbott with additional reporting by Rex Dalton and Asako Saegusa.**

On other pages | [Fall-out from nuclear cover-ups: p14](#) | [Publishing dilemmas: p15](#) | [Defining fraud: p16](#)



be significantly higher than that reported. In one survey, published in *American Scientist* in 1993, for example, between six and nine per cent of respondents said they were personally aware of results that had been plagiarized or fabricated within their faculties.

A survey in 1995 of nearly 300 randomly selected researchers in Norway revealed that 22 per cent of respondents were aware of 'serious breaches of research ethical guidelines', with nine per cent reporting that they had personally contributed to such serious breaches. Nearly 60 per cent said they were aware of less serious misconduct within their faculties. Yet the Norwegian Committee on Scientific Dishonesty has had only nine cases referred to it since it was established in 1994 — only two of which were found to have substance.

Whatever the true level of scientific misconduct, most agree that prevention is a priority. Kenneth J. Ryan chaired the Commission on Research Integrity which reviewed US policy on scientific misconduct in the mid-1990s at the request of Congress. He says that the commission's remit "was not a question of punishment, which is what most scientists worry about, but of the integrity of science — how to teach people and to prevent misconduct".

This is certainly the prevailing view in Europe, where many scientists have watched with dismay how investigations of alleged fraud in the United States have become increasingly protracted, demanding sub-

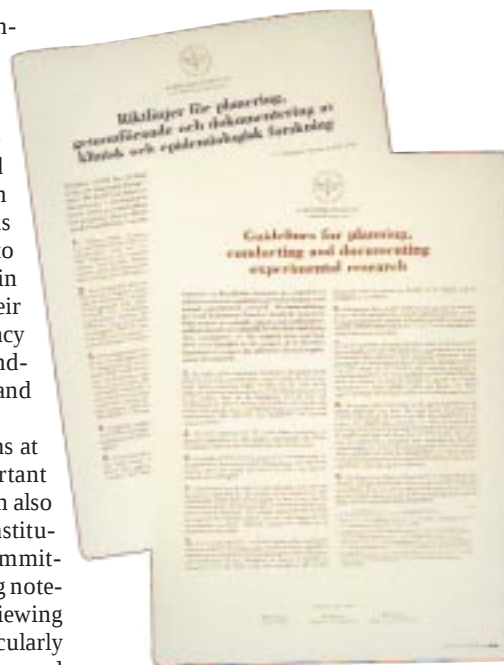
stantial amounts of time from the scientists who become involved.

#### Over-due process?

Like most organizations in Europe — except Denmark — both the NIH and the US National Science Foundation (NSF) rely on the academic institutions that receive their research funds to investigate allegations of misconduct in the first instance, and then to report their findings to the funding agency. Agency officials then review the institution's findings, investigate further if necessary, and recommend appropriate sanctions.

Keeping the primary investigations at the institutional level affords important protection to those accused. But it can also create an enormous burden for the institutions themselves. An investigating committee may spend many months analysing notebooks containing raw data and interviewing witnesses, a process than can be particularly taxing in the United States when the accused brings in an aggressive defence lawyer.

So it can take years to reach a final decision. Pre-agreed time limits are frequently waived by both parties when they become unrealistic. Lawyers for the accused may attack the integrity of the individual who makes the accusation, issue legal challenges to the investigating committee members or university administrators, or accuse whistle-blowers of inappropriate personal relationships.



**Straight and narrow: the Scandinavians were among the first to develop research guidelines.**

A lawyer defending a researcher suspected of misconduct in California once hired a private investigator to produce a psychological profile of the dean handling the investigation, based on a letter he had written to the accused, labelling the dean as paranoid.

Such tactics have persuaded some administrators and scientists that they would never again participate in such investigations. William R. Brinkley, a dean at Baylor College of Medicine in Houston, Texas, and president of the Federation of American Societies of Experimental Biology (FASEB), has observed a long-running misconduct case at his own institution (*Nature* **383**, 107; 1996). "The toll it takes is enormous," he says. "People get wrapped up in these cases for years. They have told me they couldn't afford to do it again; it destroys their careers."

The experience is having a significant impact on how fraud allegations are handled. Until recently, scientists, defence attorneys and professional societies all stressed the importance of ensuring due process for the accused — a reaction to the fact that this had not been afforded during some cases in the 1980s. But many now question whether the pendulum has swung too far the other way. They are asking whether accused researchers have been granted so much due process that institutions are deterred from investigating allegations, so facilitating a climate in which a corrupt scientist might go unpunished — or even unchallenged.

Even the current president of FASEB, which has long led the demand for maximum due process for accused scientists, acknowledges that "the duplication of the [current] process wears everyone out". "We may well have built in too much due process," says

## Japanese scandals raise public distrust

**Very few cases of scientific fraud have come to light in Japan. Some attribute this to intrinsic honesty among Japanese scientists, but others say it may merely reflect a national culture in which people turn a blind eye to issues that could give rise to social tensions.**

The hierarchical structure of university faculties, particularly within the old, conservative 'imperial' universities, and the autocratic powers of individual professors, certainly discourage scientists from blowing the whistle on dubious scientific practices.

"Whistle-blowing is rare, and those bold enough to make accusations do not walk off unscathed — they may well lose their jobs," says Masanori Kaji, a historian of science at Tokyo Institute of Technology.

In addition, young scientists cannot hope to influence suspect scientific practice inside the laboratory. "Under the *koza* system at Japanese universities, where each research group is led by a powerful professor, junior members of staff have almost no voice in the way research is carried out."

As a government white paper (policy document) on nuclear energy research pointed out last year, however, public trust in science in general has plummeted in the

past few years following a series of nuclear accidents and subsequent cover-ups, including fabrication of research data at nuclear fuel recycling plants.

The white paper said that people perceived scientists as "insular" and "unwilling to disclose or share details of their research due to preoccupation with successful results".

As a result of such incidents, scientists are under increasing pressure to take the ethical aspects of research more seriously. Nearly half of the respondents in a recent public survey conducted by Kanazawa Institute of Technology said they did not believe that scientists have higher moral values than the rest of the population. Ninety-five per cent thought scientists should take more responsibility for the way they work and how they communicate the details of their research to the public.

So, despite the apparently low incidence of scientific misconduct in Japan, broader scepticism towards ethical conduct in research is relatively high. Partly in response, the Science Council of Japan — a government advisory group — published a report last year calling for a new system to promote good scientific practice at universities. A. S.

Brinkley. "It deserves a careful look." He hopes that FASEB's new Committee on Science Policy will be able to review such issues. The committee's chairman, David Brautigan, agrees. Due process may be abused, he says. "Litigation has become the defence."

#### Europe looks for a better way

The European research community has so far avoided committing itself to such lengthy procedures — and hopes to keep things that way. In drawing up rules and procedures for handling scientific misconduct, European research organizations have tried to ensure that investigations are concluded rapidly.

In principle, the procedures, based on 'peer' investigation committees and a commitment to protect both the whistle-blower and the accused, are not radically different from those in the United States. But most European organizations try to ensure speedy resolution of investigations by limiting the number of procedural stages.

Rules introduced last year at the University of Freiburg, Germany, for example, require an informal investigation into the validity of a suspicion of dishonesty, followed, if necessary, by a formal investigation. The findings of the formal committee are passed on, with appropriate recommendations for sanctions, to the university president, who then decides on sanctions.

Albin Eser, a professor of law at the university and a director of the Max Planck Institute for Foreign and International Criminal Law, helped draw up these rules and those of the Max Planck Society, which are similarly succinct. "We deliberately avoid an appeal stage, because it would extend the time needed to reach a resolution," he says. "In any case, a sanctioned scientist can always appeal through the courts."

Scandinavian countries also leave appeals to the law courts. In contrast, the UK Medical Research Council, whose procedures are otherwise similar to those in Germany, insists on

incorporating its own appeals procedure, on the grounds of fairness.

#### National versus local

The issues of whether countries should set up national committees to investigate allegations of scientific misconduct, and whether such bodies, rather than local institutions, should be the first line for investigations, are very sensitive in Europe.

Denmark is the only country that entrusts the investigation of all allegations of scientific misconduct in the first instance to a national body, the Committee on Scientific Dishonesty, which is chaired by a high court judge. Daniel Andersen, its vice-chairman, strongly defends the centralization of investigations: "We discourage cases from being handled locally as there is a natural reluctance on the part of a university to label one of its scientists — particularly a prominent faculty member — as a cheat."

Claude Griscelli, director general of the

## Editors debate whether to blow the whistle on suspect papers

Many journal editors believe that breaches of the traditional ethics of scientific publication are increasing. But few are confident of how they should react. Should they retract a paper which only some of its authors, or a committee investigating alleged misconduct, say includes fraudulent data? Should they 'blow the whistle' on the authors of suspect submissions that they intend to reject?

There has recently been some movement on these issues. In Britain, for example, a group of particularly frustrated medical journal editors recently created the Committee on Publication Ethics (COPE) as a sort of informal self-help group. Separately, some individual journal editors are starting to develop their own policies.

Many journals, including all those published by the US National Institutes of Health, now require each author to sign a statement accepting responsibility for the whole content of a paper bearing his or her name — a move intended to avoid the perils of honorary authorship.

One significant practical test of the willingness of journals to help contain the impact of fraudulent publication is a project set up last year by the Deutsche Forschungsgemeinschaft and the Mildred Scheel Foundation, and headed by Ulf Rapp, professor of molecular and cell biology at the University of Würzburg.

The project will systematically determine how many of the 550 journal papers and 80-odd book chapters written by two German cancer researchers, Friedhelm Herrmann and Marion Brach, and some of their former colleagues, included apparently fabricated data. Local investigation committees had already identified 47 papers which appeared

to include fraudulent data. The task force, now one-fifth of the way through its stack, has identified a further 11.

Last June, Rapp approached 120 editors and publishers to request copies of relevant articles, copyright assignments and reviewers' comments. His experience, he says, was "rather depressing". Out of 70 replies, 40 sent copies of the articles, but only four provided copyright assignments, and only five reviewers' comments.

Some editors, although keen to help, did not hold files for long after publication, he said. But most refused to provide reviewers' reports, even anonymously, on the grounds of confidentiality. "Some seemed to consider us impertinent disturbers of the peace."

More than half of the 18 journals identified by the original investigation committees as likely to have published papers including fabricated data have not issued retractions. One is *Blood*, which published seven of the papers. Its editor, Kenneth Kaushansky, professor of haematology at the University of Washington School of Medicine in Seattle, says he is "uncomfortable" with this, but had been advised against retraction by *Blood's* new advisory committee on scientific integrity, on the grounds that "definitive" proof of fabrication had not been provided.

The issue raises a general dilemma for editors: if authors themselves are not willing or available to retract a paper, whose judgement that a paper includes fraudulent data should a journal accept in order to make the decision itself?

Of those editors who published retractions of the data involved in the Herrmann and Brach affair, some acted on

the requests of co-authors, others on information from one of the local investigation committees. But others remain uncertain how to react. "We need a European conference of editors to help us define exactly how, and on what basis, to retract," says Nicole Muller-Bérat, the Paris-based editor-in-chief of *Leukemia*.

*Nature's* policy is to publish whatever information it can about published papers that have proved suspect or false, says editor Philip Campbell, "although only after due consultation. Where appropriate we try to persuade the authors to issue a formal retraction, as promptly as possible. But there are times when authors will not agree among themselves. We cannot act as a judge or jury, but we can alert the community to the situation, and publish a statement by some of the authors while also making it clear where there is a disagreement."

But defining policies on retracting published papers is a relatively simple task for journal editors, compared to that of defining policies on what to do when suspicions of misconduct are raised in the minds of reviewers before publication.

Some journals, such as the *Journal of Immunology*, are willing to inform institutions of suspicions raised by papers if the authors do not give an acceptable explanation. But many editors are not convinced that their role should stretch so far.

"Journals should not act like secret police," says Magne Nylenna, editor of the *Journal of the Norwegian Medical Association*. "We consider submissions with the utmost confidentiality, and it is not clear that we should go so far as to inform institutes of [suspected] wrongdoing."

French medical research organization INSERM, would like to go further. He suggests that, at least when a local institution feels unable to handle a particular, and potentially serious, case, only an INSERM committee with members drawn from other European countries could come up with the necessary expertise and distance from local academic politics to be able to carry out a truly independent investigation. The European Science Foundation is considering whether such a committee should be set up under its aegis.

In contrast, even other Scandinavian research organizations that have adopted the Danish model, including Norway, Sweden and Finland, do not require institutions to refer cases in the first instance to the national committee. Local institutions can choose to conduct preliminary investigations.

In Norway, there were strong objections even to the concept of a national committee. Jarla Ofstad, a professor of medicine at the University of Bergen, and a former chairman of the Norwegian Medical Ethics Committee, argued strongly against the creation of the Norwegian Committee on Scientific Dishonesty because it would "create too much bureaucracy around an infrequent problem". Including a judge in each of the Scandinavian committees turns them into "amateur court-rooms", he says.

The Committee for Publication Ethics, an informal group of British medical journal editors, has long campaigned for a UK national committee to handle all incidences of misconduct, at least in medical research. This proposal is currently being discussed by a committee set up by the General Medical Council, which sets professional standards, including research standards, for clinicians.

In Germany, however, where there is strong distrust of any institution with centralized power, opposition to a national body to investigate misconduct remains strong.

#### German complacency ended

Germany is the most advanced of the non-Scandinavian European countries in instituting procedures for handling and prevent-

## US stalls on new definitions of misconduct

The US Office of Research Integrity defines misconduct as "fabrication, falsification, plagiarism ['FFP'], or other practices that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research. It does not include honest error or honest differences in interpretations or judgements of data".

Most US organizations have adopted this definition, with minor amendments or additions. Three years ago, however, the Commission on Research Integrity — which had been set up by Congress to, among other things, refine the so-called 'FFP' standard — proposed that there should be a more precise definition.

The commission suggested that a new definition would help the courts to decide complex cases. One suggestion was that the term 'elimination' should be included, referring to the deliberate omission of data considered inconvenient in ensuring the desired result.

But the proposed changes have encountered considerable hostility from scientists. Many fear, for example, that identifying such a practice as misconduct could lead to sloppy record-keeping becoming grounds for career-damaging misconduct allegations.

The recommendations are currently being tossed around the White House's Office of Science and Technology Policy, and this has given scientists another reason to worry. This is because any new definition to emerge from this office may apply government-wide, covering research funded by the National Institutes of Health, the National Science Foundation, and many other agencies involving energy, defence, education and veterans' hospitals.

Kenneth J. Ryan, a Harvard University physician and professor emeritus who chaired the Commission on Research Integrity, says that policymakers are aware of the hostility surrounding the proposals, and have therefore been in no hurry to conclude their deliberations, which have now stretched over years.

Some federal officials working on the new policies say that the slow pace is to be expected, given the importance of the issue. "It did, after all, take nearly a decade to formulate new regulations for human subjects research," says one official.

But scientists and officials involved in the work of the commission fear the importance of the issue has "dropped off the radar screen", and that it will stay there until an egregious case reaches a high enough profile to prod policymakers into action. R. D.

ing scientific misconduct. Procedures for handling allegations were drawn up by the Max Planck Society in 1997, only months before the first of a series of major fraud scandals shattered German complacency that the country was culturally immune to what had been seen as an 'American scourge'.

The Deutsche Forschungsgemeinschaft (DFG), Germany's main grants agency, was deeply shocked by the first case, when local investigation committees announced in 1997 their "strong suspicions" that 47 papers by two high-flying cancer researchers, Friedhelm Herrmann and Marion Brach, had

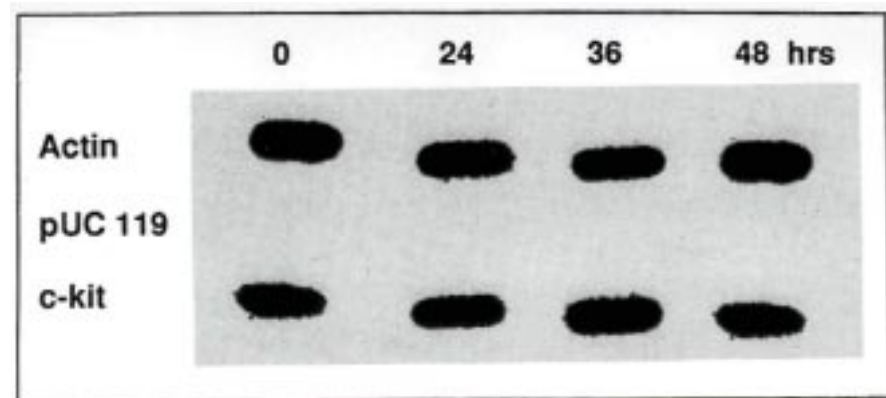
included fabricated data (see *Nature* 387, 750; 1997).

Former DFG president Wolfgang Frühwald had said with some pride in 1995 that "the incentive [for a researcher] to falsify data to accelerate his career is greater in [the US] system than in the German research system, which is tightly controlled by self-regulation". Almost as a gesture of atonement for this naive assumption, he ensured that the first guidelines for good scientific practice in Germany were rapidly developed before his term of office expired at the end of 1997.

These guidelines are intended to tighten up self-regulation rather than provide a substitute for it. Their publication prompted action by research organizations elsewhere in Europe, which had previously been considering in a more leisurely manner how to develop procedures for handling — and preventing — scientific misconduct.

In Britain, the Biotechnology and Biological Sciences Research Council launched good practice guidelines at the end of last year, and other research councils are following suit. In France, INSERM distributed guidelines to its institutions last month, and the biggest French research organization, CNRS, is currently debating the issue.

There is little to differentiate the good practice guidelines that have been developed within different research organizations. All



**Smoking gun:** investigators in the Herrmann and Brach case electronically separated the components of this autoradiogram and found it to contain duplicate data. Lane 1 (0 hrs) is the same as lane 4 (48 hrs), and lane 2 is the same data as lane 3, but part of the figure is rotated.





**Hildt: favours idea of ombudsmen.**

stress the importance of issues such as establishing clear lines of responsibility in a research team, of conducting experiments in an open way, of good lab notebooks and archiving of data, the value of mentors and the teaching of good practice to young researchers, and responsibility in publishing.

In addition, the DFG guidelines recommend the appointment of an institution ombudsman, a sort of official 'uncle' with whom young researchers could discuss in confidence their concerns about dishonesty in their laboratories. The Herrmann and Brach affair had revealed how dangerous the hierarchical culture of a German laboratory can be, since most young scientists are afraid to challenge the all-powerful figure of the professor.

But, while most researchers accept that codes of good practice are important in raising awareness, they also accept that such codes cannot in themselves eliminate misconduct: the culture of a particular laboratory, and its individuals, can always connive to ignore or undermine them.

According to Eberhardt Hildt, the whistle-blower in the Herrmann and Brach case, it would have helped if there had been an ombudsman at the Max Delbrück Centre for Molecular Medicine in Berlin, where he worked with the two researchers. It was Hildt's PhD supervisor in distant Munich who acted in this capacity and helped him to bring the affair to the attention of a disbelieving community. But, he says, good notebook keeping would not have helped.

In contrast, one young PhD student, who declines to be identified for fear of retribution, says that an ombudsman would not have helped him to expose the wrongdoing he witnessed. He was nearly driven out of research by his undergraduate experience of regular, and unexposed, manipulation of data by group leaders in a big university laboratory in Germany where he conducted his final-year project.

Young scientists in the laboratory where the young PhD student witnessed misconduct judged the risk too high to make it worth exposing what they saw as "minor misconduct" — the exaggeration of the number of controls, or the omission of data points to clean up a graph, to speed the path to publication.

He says the laboratory chief, a successful but demanding task master, is almost certainly unaware that his group leaders are massaging data, and conducting experiments that they do not report at lab meetings, in order to have results to show him in a dry period. "For

the group leaders it is not fraud or deviation from good practice: it is good tactics," says the student.

#### Enforcing good practice

The DFG predicted that most universities, which guard their independence closely, would need a serious incentive to introduce procedures for handling fraud allegations and encouraging good scientific practice. So it states in its own guidelines that institutes which do not have rules by 2002 will not be eligible to receive its funds.

The threat seems to have worked: ten out of 82 universities have brought out guidelines and, according to Josef Langer, general secretary of the German Universities Rectors Conference, the others have all agreed to do so.

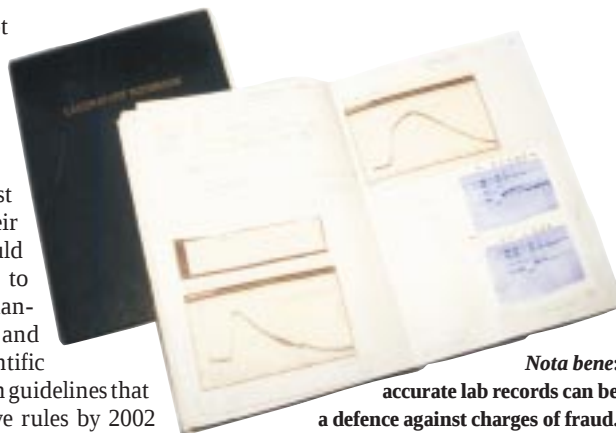
Britain's research councils have developed a less direct way of persuading universities to adopt codes of practice. By requiring grant holders to sign a statement that the work will be carried out in institutions with acceptable codes of practice, they transfer responsibility to the grant holder if misconduct occurs.

So far French universities are not being pressured to adopt codes of practice. A spokesman for the French University Presidents Conference insists that such codes are not necessary "as university science is intrinsically honest". A similar situation exists in Italy.

US research organizations operate an assurance system in grant giving, similar in principle to the new system of the UK research councils. To be eligible for grants, institutions must agree to their research being conducted and monitored according to a long list of conditions, which include the substance of the good practice codes being discussed in Europe. The enthusiasm with which good practice is enforced varies greatly between institutions, however. The NIH encourages awareness by offering training grants for young scientists with the stipulation that they must be taught research ethics.

#### Reparations

Most institutes try to ensure that papers are retracted when a scientific misconduct case is proven. But research organizations in Ger-



**Nota bene:**  
accurate lab records can be  
a defence against charges of fraud.

many in particular have made it a priority to devote substantial energy to isolating and repairing damage to the research community incurred by scientific fraud.

To disentangle truth from lies in the Herrmann and Brach affair, for example, the DFG and the Mildred Scheel Foundation — which had both given grants to the two researchers — set up an investigation to trace the history of the fabricated data, and determine if the alleged practice of fabrication had spread to other institutes from Herrmann and Brach. The investigators used software to analyse the components of data presented in the 550 papers and 80 book chapters written by Herrmann and Brach, and some former colleagues (see figure opposite). It will take 18 months to complete, and the results, detailing the corrected record, will be published in an international journal.

In a similar fashion, the Max Planck Society required its Institute for Plant Breeding in Cologne to repeat all the experiments reported in papers that had relied on assays that a technician later admitted she had manipulated (see *Nature* **393**, 293; 1998).

This task was completed within a few months with the help of researchers from within the institute and from outside. The results, and the list of papers that contained data that cannot be reproduced, are being published this month in the *Plant Journal*.

Research organizations in other European countries, as well as in the United States, have expressed admiration for this commitment to damage limitation. But none has, as yet, decided that it should do the same. □

## Guidelines on the web

The following websites include further details of guidelines on good scientific practice or guidelines for handling allegations of scientific misconduct, in English.

- US Office of Research Integrity  
ori.dhhs.gov/regguide.htm
- UK Medical Research Council  
www.mrc.ac.uk/mis\_con.pdf  
or www.mrc.ac.uk/w\_n1.html

#### ● UK Biotechnology and Biological Sciences Research Council

www.bbsrc.ac.uk/opennet/structur/hrg/sciconco.html

#### ● Germany's Max Planck Society

www.mpg.de/fehlengl.htm

#### ● Deutsche Forschungsgemeinschaft

www.dfg.de/aktuell/self\_regulation.htm

#### ● Danish Committee on Scientific Dishonesty

www.forskraad.dk/spec-udv/uvvu/



# Papers vanish in mis-citation black hole

Sir — The assessment of research quality is increasingly based on impact factors and citation analyses of published work. But there are possible biases in these measures<sup>1–4</sup>. Is the fact that Italian, French, German and Japanese publications received a less than average share of citations due to a lack of quality<sup>3</sup> or to systematic bias<sup>1</sup>?

Further analysis of the data in ref. 1 on papers by Italian scientists reveals a strong positive relationship between journal impact factor class (IFC) and proportion of publications undercited (Fig. 1). Furthermore, the proportion of publications undercited in the highest IFC is significantly greater than the corresponding proportion in the lowest IFC ( $\chi^2 = 4.89$ ,  $P = 0.027$ ). These effects strengthen the negative implications of underciting for Italian scientists: not only is their work undercited in general but, as the quality of their work increases, it is less recognized! These results clearly support Paris *et al.*'s conclusions<sup>1</sup> that undercitation does not result from substandard publications<sup>3</sup>.

What drives the undercitation? I suggest that incorrect citation may be one factor. Price<sup>2</sup> has analysed the citations of three highly cited publications, with astonishing

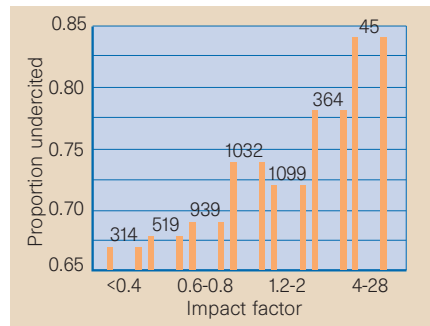


Figure 1 Relationship between the proportion of Italian publications undercited and the journal impact factor class (Spearman's  $r = 0.964$ ,  $N = 7$ ,  $P < 0.001$ ). The numbers above the bars indicate the number of publications in each class (data from ref. 1).

results: the publications were cited incorrectly in more than 200 different ways. An analysis of the data for the G7 group of countries from Table 1 in ref. 3 reveals that all non-English-speaking countries receive less than their share of citations for their share of publications, whereas the English-speaking countries all receive an equal or greater share of citations for their share of publications. This difference is statistically significant

(Fisher's exact  $P = 0.029$ ). Given the high incidence of incorrect citations overall<sup>2</sup>, there may also be a general tendency towards making more mistakes when handling citations with unfamiliar names from another language.

When this tendency is coupled with the fact that almost half of all scientific publications come from English-speaking countries<sup>3</sup>, the hazard of incorrect citation may not be random, but instead biased towards non-English-speaking countries. Because incorrect citations do not appear in citation indexes<sup>2</sup>, undercitation may be a result of incorrect citation rather than an indicator of poor-quality research in non-English-speaking countries.

Nevertheless, the reason behind the positive relationship between impact factor and undercitation (Fig. 1) remains a mystery that urgently requires further attention.

**Janne S. Kotiaho**

Department of Zoology,  
University of Western Australia,  
Nedlands, WA 6009, Australia

1. Paris, G., De Leo, G., Menozzi, P. & Gatto, M. *Nature* **396**, 210 (1998).
2. Price, N. C. *Nature* **395**, 538 (1998).
3. May, R. M. *Science* **275**, 793–796 (1997).
4. Barreto, G. R. *Science* **276**, 882–883 (1997).

## Engineering a longer life

Sir — In the excellent article "Eighteen-ninety-nine and all that", Isambard Kingdom Brunel is said to have died in 1869 (*Nature* **397**, 15–18; 1999). Would that he had! We would have had ten years more of one of the world's greatest engineers.

**Nigel L. Firth**

Nova Scotia Agricultural College,  
Truro, Nova Scotia B2N 5E3, Canada

## Managing SOHO

Sir — Tony Reichhardt's News article "Rescued satellite to get more managers" is almost entirely accurate in its explanation of the concerns of operating a complex spacecraft during a period of pressure to reduce the costs of flight operations (*Nature* **396**, 399; 1998). But it is misleading in stating that "Management of the \$1 billion satellite was to have merged with that of the other projects in the ... ISTP programme to save money".

In fact, it was the operation of SOHO and the ISTP WIND and POLAR satellites that was to have been merged in order to reduce costs, and indeed this is still going ahead. The management of SOHO flight operations, resources allocation and scientific investigations was never planned to be merged with the rest of ISTP. Indeed, the plan was to continue the already existing, shared management of SOHO, WIND, POLAR and several other space science assets operated at NASA Goddard.

I believe the new SOHO programme office offers an opportunity to manage a major space science asset in a manner consistent with both its scientific importance and the level of operational risk.

**Joseph B. Gurman**

NASA Goddard Space Flight Center,  
Code 682, Greenbelt, Maryland 20771, USA

## Controversy over the cloning of mice

Sir — Tsunoda and Kato<sup>1</sup> report the successful cloning of mice after transferring cell nuclei from late preimplantation

embryos. Our earlier studies<sup>2</sup>, using different methods, were mostly neglected or rejected. But Tsunoda and Kato found similar, although not identical, results concerning the ability of nuclei from the inner cell mass (ICM) of blastocysts to give rise to adult mice.

It is important to note that, under different experimental conditions, similar results have confirmed the developmental potential of ICM nuclei. Tsunoda and Kato also report the successful cloning of mice after transferring cell nuclei from the trophoblast of blastocysts, a type of cell thought to be specialized for its extra-embryonic development.

In contrast, McGrath and Solter<sup>3</sup> concluded from negative results that the cloning of mice using donor nuclei from late preimplantation embryos was impossible, and suggested that the "inability of cell nuclei from these stages to support development reflects rapid loss of totipotency of the transferred nucleus".

Now, with these positive data on the cloning of mice, the time has come for the correct evaluation of our earlier results on the first cloning of a mammal.

It has been argued that the early activation of the mouse embryonic genome is a barrier to cloning, leaving too little time

for the nucleus to be reprogrammed. But even Solter<sup>4</sup> now admits that this is not insurmountable, in the light of successful experiments by Wakayama *et al.*<sup>5</sup> using adult cumulus cells for the cloning of mice.

**Karl Illmensee**

Universitätsklinik für Frauenheilkunde,  
Anichstrasse 35, A-6020 Innsbruck, Austria

1. Tsunoda, Y. & Kato, Y. *J. Reprod. Fertil.* **113**, 181–184 (1998).
2. Illmensee, K. & Hoppe, P. *Cell* **23**, 9–18 (1981).
3. McGrath, J. & Solter, D. *Science* **226**, 1317–1319 (1984).
4. Solter, D. *Nature* **394**, 315–316 (1998).
5. Wakayama, T., Perry, A. C. F., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. *Nature* **394**, 369–374 (1998).

## Forging links in an electronic paper chain

Sir — The Briefing on electronic journals was interesting and timely (*Nature* **397**, 195–200; 1999). But one issue that was not addressed was how Internet publishing will change the style of scientific writing. One imagines that the length of on-line articles will be less restricted than paper ones — even in the most selective journals. This will encourage a more thorough, but perhaps windier, writing style.

Counterbalancing this are the possibilities of hypertext. This will allow authors to connect their articles to supplementary material on their own sites or in external databases. This will enable them to reduce the main body of their text and to make it less technical, moving the details to linked sections. It may also lead to a more segmented, 'fact-box' style of presentation. Copious links will require careful layout, ensuring that they remain stable and reflect an underlying logic.

Finally, the use of hypertext in papers raises the issue of whether authors will be free to modify linked material on their own websites, or whether the content related to a paper should be frozen on submission. This is especially relevant to the refereeing process.

**Mark Gerstein**

Department of Molecular Biophysics and Biochemistry, Yale University, Bass 432A,  
266 Whitney Avenue, New Haven,  
Connecticut 06520-8114, USA

## The editor as an endangered species

Sir — I used to feel a great sense of security in my job as editor of *Physical Review Letters*. Receipts continue to increase, the journals of the American Physical Society are leading in most aspects of electronic publishing, and, of course, an editor could

never be replaced by a computer program.

Alas, your Briefing on electronic journals tells me I am an endangered species (*Nature* **397**, 195–200; 1999). Apparently the *Journal of High Energy Physics* already has a robot that reads manuscripts and assigns them to referees. I imagine your reporter meant to say "assigns them to editors". And, according to a picture in your article, my boss Martin Blume has become a web page! He was spotted in the editor-in-chief's office recently, so that must have been a printer's error.

No doubt the electronic future will have robots that will avoid such errors, but will readers be able to trust them?

Please tell me that editors are really needed.

**Gene L. Wells**

Physical Review Letters,  
1 Research Road, Box 9000, Ridge,  
New York 11961-9000, USA

## Space-grown crystals may prove their worth

Sir — The first building block of the International Space Station was launched on 20 November 1998, but the potential uses of the space station are still under debate. A recommendation to scrap NASA's research on protein crystals was reported recently<sup>1</sup>. The reason given was that no serious contributions to our knowledge of protein structure have yet been made in space. We wish to point out, on the basis of recent experimental and theoretical evidence, that in many cases the potential benefits of the microgravity environment have not been fully exploited. This explains the low rate of success of protein crystallization in microgravity and opens up the scope for enhancing the efficiency of experimentation in space.

Microgravity eliminates sedimentation and convective mixing, so offering a more homogeneous growth medium compared with growth on Earth. Since this is likely to improve the degree of perfection of the crystals, why has microgravity crystallization not been more successful?

There are four common methods for crystallizing proteins: batch, vapour diffusion, dialysis and free interface diffusion (FID). Vapour diffusion is the most successful technique for crystallization on Earth. Naturally it became the method of choice for crystallization in microgravity. Thanks to the European Space Agency providing new means of conducting experiments in a far more systematic way, a comparison of microgravity crystallization using different

methods was facilitated. The results demonstrated that vapour diffusion is not the best technique for crystallization in microgravity<sup>2</sup>.

Images from CCD cameras recorded during flights showed that some crystals grown by vapour diffusion displayed a cyclic motion within the aqueous drop in which they grow<sup>3</sup>. This motion is attributed to Marangoni convection, an effect which serves to reduce concentration gradients along the interface between the solution and the vapour<sup>4</sup>. In the case of FID and dialysis there is no interface between solution and vapour and this cyclic motion does not occur<sup>5</sup>.

Cyclic movement of the crystals in microgravity destroys the very benefit that is sought from the unique environment of outer space and thus may be a limiting factor in the ultimate perfection (indicated by X-ray diffraction) of the crystals that can be obtained<sup>6–8</sup>.

Several researchers have mentioned that crystals grown in microgravity by dialysis and FID methods appeared to be superior to those grown by vapour diffusion<sup>9,10</sup>, but those results were not taken seriously enough and most experiments were still done by vapour diffusion. Recent video recordings<sup>3,5</sup> show beyond any doubt that crystal movement (akin to sedimentation referred to above) takes place in the case of vapour diffusion but not with the other methods.

It is apparent that we may have only now grasped how best to use the microgravity environment. Hence it would be a great shame if the experiments were scrapped now, just at the stage when a better understanding of crystallization in space and its fluid physics and biophysical chemistry is being gained. We are now in a position to explore more efficient ways of increasing the success rate of these experiments. The translation of the results to improve protein structure determination will come later. The stage of basic research is not yet completed to allow targeted exploitation to take place.

**Naomi E. Chayen**

Physics Department, Imperial College of Science,  
Technology and Medicine, London SW7 2BZ, UK

**John R. Helliwell**

Department of Chemistry, University of Manchester,  
Manchester M13 9PL, UK

1. Reichhardt, T. *Nature* **394**, 213 (1998).
2. Boggan, T. J. *et al. Phil. Trans. R. Soc. Lond. A* **356**, 1–17 (1998).
3. Chayen, N. E. *et al. J. Cryst. Growth* **171**, 219–225 (1997).
4. Molenkamp, T., Janssen, P. B. M. & Drenth, J. "Protein crystallisation and Marangoni convection. Final reports of sounding rocket experiments in fluid science and materials science" ESA Report SP-1132, 22–43 (ESA, Paris, 1994).
5. Snell, E. H. *et al. Acta Crystallogr. D* **53**, 747–755 (1997).
6. Snell, E. H. *et al. Acta Crystallogr. D* **51**, 1099–1102 (1995).
7. Normile, D. *Science* **270**, 1921–1922 (1995).
8. Savino, R. & Monti, R. *J. Cryst. Growth* **165**, 308–318 (1996).
9. Koszelak, S. *et al. Biophys. J.* **69**, 13–19 (1995).
10. Eposito, L. *et al. Acta Crystallogr. D* **54**, 386–390 (1998).

# From rifting to drifting

Fred F. Pollitz

Africa is splitting apart down the East African Rift. An analysis of the relative motion of the tectonic plates involved allows the rate of continental rifting to be estimated.

Geologists are naturally drawn to modern examples of fundamental processes that have occurred over long stretches of Earth's history. One such process resulted in the formation of the world's mid-ocean ridge system — a 60,000-km-long suboceanic mountain chain, which rises about 2 km above the surrounding ocean floor, and which was created in progressive stages by break-up of previously intact continents<sup>1</sup>.

A modern equivalent of an incipient mid-ocean system is the East African Rift (Fig. 1), a region of continental break-up, 4,000 km long, that stretches from Djibouti to Mozambique. Diagnostic indicators such as seismicity and recent volcanism have long given hints that this rift system is actively opening. In a new interpretation of marine geophysical data, Chu and Gordon (page 64 of this issue<sup>2</sup>) provide quantitative verification that this is indeed so, and they show that the northern part of the rift is opening at up to 6 mm per year — that is, at about half the speed of a 'slow' oceanic spreading centre.

The basic tenet of the theory of plate tectonics is that the 50–200-km-thick outer shell of the Earth can be divided into a number of large blocks or plates. The instantaneous relative motion between any two of them is, in principle, completely described in terms of rigid rotation about a fixed pole termed the Euler pole. If one plate is regarded as fixed, then the moving plate sweeps out a path along small latitudinal circles centred on this pole at rates of between about 10 and 120 mm per year. Plate tectonic theory has been tested not only with geological reconstructions of continental margins now separated by vast oceans, with supporting evidence from the distribution of sediments on the ocean floor, but also by using purely geometrical measurements. Such measurements include seafloor-spreading rates deduced from the seafloor magnetic anomalies produced by a combination of spreading and reversals of the Earth's magnetic field; the orientation of oceanic fracture zones which parallel the small circles; and the relative slip direction along a plate boundary, as inferred from earthquakes.

In the 1960s, many data of these various types were brought to bear on delineating the relative motions across the global plate

boundaries, particularly the oceanic rift system. The fact that a set of relative rotation poles could be constructed which satisfies the constraints imposed by these data, and completes a circuit around the globe, was a remarkable demonstration of the predictive power of plate-tectonic theory.

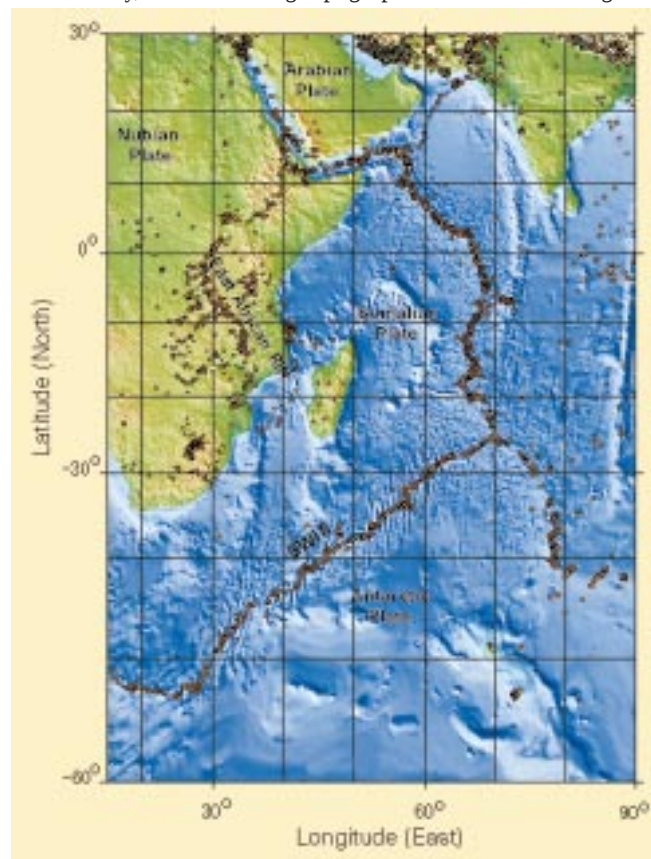
Later refinements of this model have consisted essentially of revising and expanding the set of geometrical measurements and recognizing that continental and (to a lesser extent) oceanic interiors may exhibit significant non-rigid behaviour<sup>3–6</sup>. Chu and Gordon<sup>2</sup> now contribute by presenting compelling evidence that the African plate consists of two blocks, the Nubian (West Africa) and Somalian (East Africa) plates. These blocks are themselves rigid, but are splitting apart along the East African Rift, at predictable rates, and so confer non-rigidity on the African plate as a whole. Geological evidence — in the form of a corresponding belt of seismicity, and its striking topographical

similarity to the actively opening mid-ocean ridges — had already pointed to this conclusion; but no geometrical demonstration comparable with the earlier tests of global plate tectonic theory had been successfully carried out.

The key ingredient in Chu and Gordon's analysis is the recognition of systematic differences in the rate and direction of Africa-to-Antarctica relative motion along the Southwest Indian Ridge (SWIR; Fig. 1). Most apparently, fracture zone orientations on the SWIR exhibit a clockwise bias on the Nubian side, and spreading rates are low on the Somalian side, compared with the orientations and rates expected if Africa behaved as a single rigid plate.

Counterintuitively, the extra Nubia–Antarctica relative motion on the SWIR is towards the east, and the extra Somalia–Antarctica motion is towards the southwest, in apparent conflict with the rifting apart of Nubia and Somalia within the African continent. Chu and Gordon resolve this puzzle by placing the Nubia–Somalia Euler pole well to the north of the SWIR, and yet well to the south of the centres of continental rifting (see their Fig. 2 on page 65). Their model is consistent with the broad range of possible Nubia–Somalia relative motions inferred from closure of the Arabia–Nubia–Somalia plate circuit.

Earth scientists have, of course, thought quite deeply about the origin of the East African Rift without waiting for proof that the bounding blocks are actually drifting



**Figure 1 Seismicity of the Indian Ocean and adjacent landmasses, 1975–95, showing the East African Rift. Also shown are the Southwest Indian Ridge (SWIR), and the Nubian, Somalian, Arabian and Antarctic tectonic plates. Chu and Gordon's<sup>2</sup> estimates of the opening rate of the East African Rift are 3 mm per year at 15° S and 6 mm per year at 10° N. (Map courtesy of the USGS National Earthquake Information Center; see ref. 13.)**



apart. With direct evidence now at hand for the tangible drifting apart of East and West Africa, attention may turn again to the forces behind it. Rifts are often categorized as active or passive depending on whether the forces driving the bounding blocks apart are local or distant in nature. Several lines of evidence suggest that an upwardly mobile plume of hot material originating very deep beneath northeast Africa provides the huge forces required to actively split the continent<sup>7,8</sup>. Such a 'hotspot' beneath an oceanic plate tends to penetrate the plate and dissipate its thermal energy in bursts of volcanism which eventually produce linear island chains such as the Hawaiian islands.

An old, geologically inactive continental interior, however, such as that which existed in northeast Africa before 40 million years (Myr) ago, is colder and more difficult to penetrate, leading to stagnation of an incident plume head at the base of the continental plate. During the millions of years needed to heat up and weaken the deeper continental plate, this plume head must spread out horizontally and exert enormous drag forces on the surrounding plate or plates. Gravitational forces accompanying the simultaneous uplift of the region can further increase the forces radiating outward from the plume centre.

An acceleration in East African Rift opening rate about 6 Myr ago is implied by the opening rates of up to 6 mm per year in Chu and Gordon's model. With a 35–40-km void left behind by the opening within the northern rift zone, these rates cannot easily be extrapolated back to the time of incipient rifting around 30 Myr ago but only back to about 6 Myr, a time that coincides with substantial changes in both African and Arabian plate motions<sup>9</sup>. A large plume interacting with northeast Africa starting about 40 Myr ago probably triggered the initial period of slow rifting and volcanism, and appears ideally situated to have accelerated the break-up. More work will be needed to understand the complex history of interaction between the rising plume and the African plate, and to explain how and why their mechanical interaction accelerated relatively recently. In any event, if present opening rates continue unabated, the northern part of the rift will probably resemble the Red Sea in about 10 Myr.

A long-term benefit of Chu and Gordon's work<sup>2</sup> will be improvement in the accuracy of global models of instantaneous plate motion. As DeMets *et al.*<sup>10</sup> have pointed out, failing to split a plate into two or more independently moving parts or, to the same effect, inaccurate estimation of relative plate motion along a particular plate boundary, leads to systematic errors in predicting relative plate motions along all of the other recognized global plate boundaries. Although further revisions exceeding 10 mm per year would be surprising, revisions of Indo-

Australian<sup>10,11</sup>, Caribbean<sup>12</sup> and African<sup>2</sup> plate motions approaching that magnitude are a reminder that the big picture needs constant scrutiny.

Fred F. Pollitz is in the Department of Geology, 1 Shields Avenue, University of California, Davis, California 95616, USA.

e-mail: pollitz@geology.ucdavis.edu

1. Heirtzler, J. R. *et al.* *J. Geophys. Res.* **73**, 2119–2136 (1968).
2. Chu, D. & Gordon, R. G. *Nature* **398**, 64–67 (1999).
3. Atwater, T. *Geol. Soc. Am. Bull.* **81**, 3513–3536 (1970).

4. Molnar, P. & Tapponnier, P. *Science* **189**, 419–426 (1975).
5. McKenzie, D. P., Davies, D. & Molnar, P. *Nature* **226**, 243–248 (1970).
6. Weissel, J. K., Anderson, R. N. & Geller, C. A. *Nature* **287**, 284–291 (1980).
7. Ebinger, C. J., Bechtel, T. D., Forsyth, D. W. & Bowin, C. O. *J. Geophys. Res.* **94**, 2883–2901 (1989).
8. Zeyen, H. *et al.* *Tectonophysics* **278**, 329–352 (1997).
9. Pollitz, F. F. *Tectonophysics* **194**, 91–106 (1991).
10. DeMets, C., Gordon, R. G., Argus, D. & Stein, S. *Geophys. J. Int.* **101**, 425–478 (1990).
11. Royer, J.-Y. & Gordon, R. G. *Science* **277**, 1268–1274 (1997).
12. Dixon, T. H. *et al.* *J. Geophys. Res.* **103**, 15157–15182 (1998).
13. <http://www.neic.cr.usgs.gov/neis/general/seismicity/indian.html>

## Protein recognition

# An SH2 domain in disguise

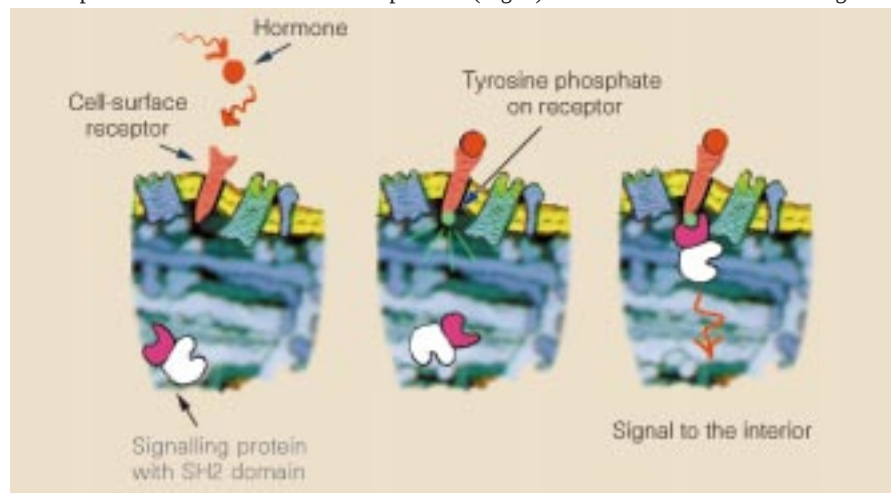
John Kuriyan and James E. Darnell Jr

The evolution of multicellular animals appears to be coincident with the development of a unique signalling device known as the Src-homology-2 (SH2) domain<sup>1</sup>. So named because of their original discovery in the genomes of sarcoma viruses, these protein modules are dedicated to the transmission of molecular signals that start at the cell surface and then engage the inner workings of the cell<sup>2</sup>. SH2 domains have so far been found only in animals, and are apparently absent in fungi and plants, the other major divisions of the crown of the eukaryotic tree. Until now, we derived our ideas about the distribution and evolution of SH2 domains from their recognizable presence in the amino-acid sequences of proteins encoded by animal genomes.

On page 84 of this issue, Meng *et al.*<sup>3</sup> report the three-dimensional structure of a signalling protein from humans known as Cbl, and find that it contains an SH2 domain that is so divergent in amino-acid sequence that its presence in Cbl had not been suspect-

ed before. The identification of this distantly related but genuine SH2 domain suggests that SH2 domains are more broadly distributed in nature than their current count would indicate, and raises anew the question of where these crucial signalling domains originated.

Each cell in a multicellular animal responds to instructions and requests from other cells in the animal, transmitted in the form of secreted hormones or by proteins at the cell surface. A characteristic feature of the cellular response to these signals is the use of tyrosine phosphorylation as the crucial molecular switch in the very first steps of information transfer from the cell surface to the interior<sup>4</sup>. This involves the addition of phosphate groups to specific tyrosine residues, first on the cytoplasmic face of transmembrane receptors and thence to associated proteins in the cytoplasm and in the cell nucleus, a modification that is catalysed by enzymes known as tyrosine kinases (Fig. 1). SH2 domains act as switching units



**Figure 1** Simplified representation of how SH2 domains work. A cross-sectional view of an animal cell is shown, with the cell membrane depicted in yellow. A hormone is bound by its specific transmembrane receptor (both in red). Activation of the receptor causes phosphorylation of tyrosine (phosphotyrosine in green), which enables a signalling protein containing an SH2 domain (purple) to recognize the receptor. The signalling protein contains a catalytic domain (white), which transmits a signal when the SH2 domain is engaged. The SH2-phosphotyrosine interaction allows specific molecular pairings to be established in the crowded environment of the cell.

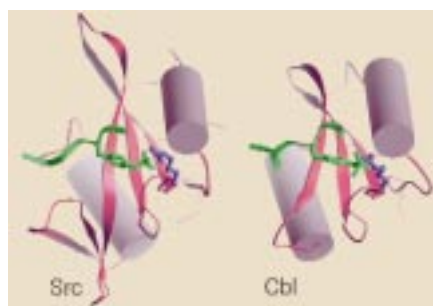


Figure 2 A 'mainstream' SH2 domain, from the tyrosine kinase of the Rous sarcoma virus (Src, left)<sup>11</sup>, is compared to that from Cbl (right)<sup>3</sup>. The backbone structure of the SH2 domain is shown in purple in both. Each of the SH2 domains is bound to a piece of protein containing a phosphorylated tyrosine residue (green). An arginine residue, conserved in all SH2 domains, interacts with the phosphate group. All the crucial aspects of phosphotyrosine recognition are similar in both SH2 domains.

by recognizing phosphotyrosine residues specifically, and phosphotyrosines serve as beacons that attract SH2-containing protein molecules. These proteins then transmit molecular signals to key cytoplasmic proteins and genes in the nucleus, thus regulating the machinery of the cell<sup>5,6</sup>.

The discovery of an SH2-containing signalling molecule in the slime mould *Dictyostelium discoideum*, along with the apparent absence of SH2 domains in yeast and in plants, has suggested a role for SH2 domains in the evolutionary transition to multicellular animals<sup>7</sup>. *Dictyostelium*, an organism that can become multicellular, spends much of its life cycle in a unicellular form. In response to environmental cues such as starvation, individual cells begin to clump together and form a 'slug', which can move as one unit. Soon, a developmental process begins. Some of the cells differentiate into a stalk, while the more fortunate ones develop into a fruiting body capable of dispersing spores. In this most simple of developmental decisions, a signalling protein known as STAT (for 'signal transducer and activator of transcription') is crucial for the determination of cell fate<sup>7</sup>. Although the full complexity of metazoan signalling is absent in *Dictyostelium*, the STAT protein in the slime mould is clearly related in its amino-acid sequence to STAT proteins in more complex organisms, including humans.

What is significant to the present discussion, however, is that both the human and *Dictyostelium* STAT proteins contain an SH2 domain, which allows the STAT proteins to dimerize in response to activating signals. They dimerize and thus become activated through reciprocal phosphotyrosine-SH2 interactions; the activated dimer can now bind to DNA and direct transcription. The SH2 domain in the *Dictyostelium* STAT is evolutionarily the most ancient such

sequence known so far, and the slime moulds are the simplest organisms yet discovered to contain an SH2-phosphotyrosine signalling protein. It may be that the *Dictyostelium* STAT (or a STAT in some other similar ancient organism) represents an evolutionary link between single-cell and multi-cell organisms.

So what about Cbl? Cbl is a so-called 'adaptor' protein which links different signalling molecules together upon activation of the receptor. Cbl was known to bind to phosphorylated tyrosine residues, but nothing in its amino-acid sequence suggested that it harboured an SH2 domain. The three-dimensional structural analysis presented by Meng *et al.*<sup>3</sup> shows that the phosphotyrosine-recognition property of Cbl is conferred by a structural unit that resembles classical SH2 domains in every important way. For example, all SH2 domains have an invariant arginine residue that rises up from an interior location in the domain to engage the phosphate group of the phosphotyrosine<sup>8</sup>. This crucial determinant of specificity is conserved in a precise structural context in the Cbl SH2 domain (Fig. 2). Until now, the STAT SH2 domains have been among the most divergent of the known SH2 domains in terms of their amino-acid sequences. The discovery of a perfectly formed SH2 domain in Cbl indicates that the range of sequences that will form SH2 domains has an even greater spread than had been suspected.

Given the lack of sequence similarity between the Cbl SH2 domain and the more familiar ones, could the structure of the Cbl domain have arisen by convergent evolution from an unrelated ancestral protein? Although this is a possibility, it is highly unlikely. The strong conservation of three-dimensional structure between Cbl and other SH2 domains suggests to us that they have evolved from a common ancestral protein. There are examples of phosphotyrosine-binding domains that are structurally unrelated to the SH2 domain, demonstrating that the need to recognize phosphotyrosine does not place any particular constraints on protein structure<sup>8</sup>.

In this regard, another intriguing observation is the presence in Cbl of a small helical domain located just before the SH2 domain. Strikingly, the STAT proteins also have a domain with a very similar structure located before their SH2 domains<sup>9,10</sup>. Although the helical domains are rotated with respect to one another in Cbl and the STATs, it is a tantalizing thought that their similar folding topology may reflect a vestigial structural remnant of an as-yet-unknown ancestral SH2-containing protein.

The identification of very distant relationships between proteins proceeds most reliably when three-dimensional information is available for representative members



#### 100 YEARS AGO

In the history of vegetable physiology, sufficient importance has not been given to Dante's observations upon the action of solar light and heat upon plants, and to the ideas upon this action that existed in Italy in the fourteenth century. ... It is not unlikely that the verses of Dante influenced Leonardo da Vinci in believing that "the sun giveth spirit and life to plants, and the soil with its moisture nourisheth them," leading him to an experiment in which the importance of leaf-function in the nourishment of plants is first noted, two hundred years before Malpighi. In this experiment Leonardo caused a water-fed plant to grow prosperously and bear fruit abundantly, although its roots had purposely been reduced to "only one tiny rootlet" (*solamente una minima radice*). Leonardo thus succeeded in causing a plant to grow chiefly by its foliage, to "*vivere della cima*" ("Paradiso", xviii. 29): an experiment that would have been too dangerous for the experimenter in Dante's days.

From *Nature* 2 March 1899.

#### 50 YEARS AGO

Last August scientific workers from all over the world heard with deep disappointment that the Soviet Union had officially adopted an isolationist attitude on certain branches of biology. For the first time in the U.S.S.R. there was established a 'party line' in one of the natural sciences. Since then there has been speculation as to whether this attitude might extend to other natural sciences, and a recent broadcast from Moscow gives point to these speculations. On January 26, 1949, the philosopher Alexander Alexandrovitch Maximov, who is a corresponding member of the Academy of Sciences, and who belongs to the staff of its Institute of Philosophy, gave a broadcast on the Moscow Radio Home Service. The theme of his talk was the correct Bolshevik attitude to natural science. He attacks those foreign physicists who "regard as synonymous the philosophical definition of matter and the objective idea of reality", and who are responsible for other "idealistic misinterpretations" in relativity and quantum theory. He indicts by name Einstein, Niels Bohr and Heisenberg.

From *Nature* 5 March 1949.

of the family. By extending the sequence range of the SH2 domains, the Cbl structure should now allow structural analysts to explore the rapidly growing genomic databases for other remote offspring of the SH2 domains. The *Dictyostelium* SH2 domain will then not be the only indicator of how the intimate relationship between phosphotyrosine and SH2 domains first got going. □

John Kuriyan is at the Howard Hughes Medical Institute, and James E. Darnell Jr is in the Laboratory of Molecular Cell Biology, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.  
e-mail: kuriyan@rockvax.rockefeller.edu

1. Darnell, J. E. Jr *Proc. Natl Acad. Sci. USA* **94**, 11767–11769 (1997).
2. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252**, 668–674 (1991).
3. Meng, W., Sawadkiosol, S., Burakoff, S. J. & Eck, M. J. *Nature* **398**, 84–90 (1999).
4. Hunter, T. & Cooper, J. A. *Annu. Rev. Biochem.* **54**, 897–930 (1985).
5. Pawson, T. *Nature* **373**, 573–580 (1995).
6. Cohen, G. B., Ren, R. & Baltimore, D. *Cell* **80**, 237–248 (1995).
7. Kawata, T. *et al.* *Cell* **89**, 909–916 (1997).
8. Kuriyan, J. & Cowburn, D. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288 (1997).
9. Chen, X. *et al.* *Cell* **93**, 827–839 (1998).
10. Becker, S., Groner, B. & Müller, C. W. *Nature* **394**, 145–151 (1998).
11. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. *Cell* **72**, 779–790 (1993).

## Cosmology

# Evolution of the cosmological constant

P. J. E. Peebles

Contrary to expectation, the evidence is that the Universe is expanding at about twice escape velocity — the speed required to overcome the gravitational pull of all the matter in the Universe<sup>1</sup>. This would mean we live at a special time: in the past, the greater density of mass in the Universe was gravitationally slowing the expansion; in the future, the expansion rate will be close to constant or maybe speeding up under the influence of a new type of matter that some call quintessence. In *Physical Review Letters*, Steinhardt and colleagues<sup>2</sup> argue that the gravitational effect of a particular form of quintessence takes a long time to grow comparable to the effect of ordinary matter and radiation. If so, it is not so surprising that we have come on the scene just as

quintessence has become a factor in the evolution of the Universe.

Quintessence began as Einstein's cosmological constant,  $\Lambda$ . It has negative gravitational mass: its gravity pushes things apart. Einstein hoped the repulsion would balance the gravitational attraction produced by ordinary matter, so the Universe could be static. The discovery of the expansion of the Universe led him to abandon  $\Lambda$  as an unnecessary hypothesis. Particle physicists later adopted Einstein's  $\Lambda$  as a good model for the gravitational effect of the active vacuum of quantum physics, although they are at odds with the small value of  $\Lambda$  indicated by cosmology. Theoretical cosmologists noted that as the Universe expands and cools  $\Lambda$  tends to decrease. When water freezes, it releases

latent heat; ice has a lower energy density than water. As the Universe cools, symmetries among forces are broken, particles acquire masses, and these processes tend to release the analogue of latent heat. The vacuum energy density accordingly decreases, and with it the value of  $\Lambda$ . Perhaps an enormous  $\Lambda$  drove an early, rapid expansion that smoothed the primeval chaos to make the near uniform Universe we see today, requiring  $\Lambda$  to decrease over time to its current level. This is the cosmological inflation concept.

A decade ago we began debating the idea that  $\Lambda$  is still present and that it is not, in fact, constant, but still decreasing. Perhaps  $\Lambda$  has a small value now simply because the Universe is old<sup>3</sup>. Steinhardt and colleagues<sup>2</sup> add the idea that in simple theoretical models of  $\Lambda$ -like matter, or quintessence, a broad range of initial conditions in the early Universe evolve to a similar situation, in which the gravitational effect of quintessence is less than that of ordinary matter, and remains that way for a long time. This means  $\Lambda$  would be expected to become important again only late in the career of the Universe, maybe about now. A more convincing story will require a better understanding of how  $\Lambda$ -like matter might fit in the fundamental basis for physics, perhaps in the dynamics of the quantum-mechanical vacuum or ultra-low energy excitations of fields we have yet to discover.

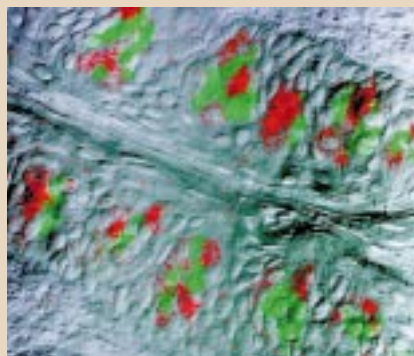
Observational cosmologists can measure the present value of  $\Lambda$ <sup>4,5</sup>. Distant galaxies are seen as they were in the past, because of the time it takes light to travel. Their recession velocities, inferred from the Doppler shifts of their light, are a measure of how fast the Universe was expanding then. The observations suggest that  $\Lambda$  does have an effect: the expan-

## Signal transduction

# A taste of things to come

It used to be the case that, if you wanted to know anything about taste, the best person to ask was probably a chef. We still know relatively little about the molecular basis of taste perception compared with, say, vision or touch. But a report by Charles Zuker, Nick Ryba and colleagues (*Cell* **96**, 541–551; 1999) now adds to the picture. Using standard molecular biological recipes, these authors have discovered two potential mammalian taste receptors.

Mammals are thought to have five basic taste modalities — sweet, bitter, salty, sour and umami (the taste of monosodium glutamate). Different regions of the tongue prefer these various modalities. So, for example, the circumvallate papillae at the back of the tongue are particularly sensitive to bitter



substances, whereas the fungiform papillae at the front of the tongue prefer sweet compounds.

Zuker, Ryba and co-workers identified the new proteins — known as TR1 and TR2 — starting from complementary DNA libraries of sequences expressed only

in taste cells. Both proteins are guanine-nucleotide-binding (G) protein-coupled receptors with sequence homologies to other candidate chemosensory receptors.

When they studied the topographic expression of TR1 and TR2, the authors found that TR1 is expressed in all fungiform taste buds, and that TR2 localizes to the circumvallate taste buds. What's more, as the image on the left shows, these proteins (green) do not colocalize with gustducin (red).

Studies using knockout mice indicate that gustducin is involved in bitter and sweet transduction. But this image shows that it is probably not responsible for signalling from TR1 and TR2. Finding the agents that are could well be the next chapter in the taste story.

Alison Mitchell



sion rate started increasing when the Universe was 60% of its present size. That can be compared to 75% of the present size when our Solar System formed.

The measurements that provide evidence for this  $\Lambda$ -like matter have been beautifully done, but there is no guarantee that all systematic errors are under control. You would do well to bet no more than two dollars to get one that  $\Lambda$  has really been detected. I am offering ten dollars to get one on the proposition, which has been more thoroughly cross-checked, that our Universe is expanding faster than escape velocity. In either case, the origin of life on Earth coincided with another cosmological event, the end of the gravitational brake on the expansion of the Universe.

There are checks on the measurement. For example, the oldest observed stars have to be younger than the time it took the Universe to expand from densities too high for stars to have existed. If the expansion were now accelerating (that is, the expansion rate was slower in the past), then the time elapsed since the Big Bang would have been longer and the Universe would be considerably older, than if the expansion were still slowing down. People are making impressive progress in the difficult art of measuring star ages. If this condition on the age of the Universe and other cosmological tests consistently indicate that the expansion is accelerating, it will compel acceptance of  $\Lambda$ .

If  $\Lambda$  has been detected, it became a serious factor in the expansion rate just as life appeared on Earth: this is known as one of the cosmic coincidences. Deciding whether a coincidence is of some significance or only an accident is not easy. If observers existed when the Universe was half its present size, they would have lived through an abrupt reduction in star formation — from a rate that had held nearly constant well into the past, to the decline that persists to the present day. Similar beings to us could have existed when the Universe was only 3% of its present size, if their Cambrian explosion had happened a lot faster than ours. Such observers may have noticed the remarkable coincidence that the Universe was too hot for water-based life just a factor of three earlier in its expansion history. Life as we know it is still possible when the Universe reaches ten times its present size, near stars formed from the dregs of interstellar gas. These observers would find the Universe a lonely place; other host stars would be exceedingly rare. And they may well find it was too late to search for life on planets in other galaxies; they could see distant galaxies, but  $\Lambda$  would be pushing them away so fast that light could no longer reach distant stars even in principle.

The great advances in detectors, telescopes and observatories on the ground and in space have given us a rough picture of what happened as our Universe evolved from a

dense, hot and maybe quite simple early state to its present complexity, and observations in progress are filling in the details. That in turn is driving intense debate on how the behaviour of our Universe can be understood within fundamental physics. Those who work on the fundamentals are well aware of the issue of  $\Lambda$ -like matter; they're certainly trying to solve this theoretical detail. It will help when we know for sure whether  $\Lambda$  has really been detected, and it would help even more if the measurements showed  $\Lambda$  to be decreasing. Maybe we will discover that there is some cosmic signifi-

cance to when we appeared — apart from our need for a comfortable planet in stable orbit around a long-lived star. □

P. J. E. Peebles is in the Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: [pjep@pupppg.princeton.edu](mailto:pjep@pupppg.princeton.edu)

1. Bahcall, N. A. & Fan, X. *Proc. Natl Acad. Sci. USA* **95**, 5956–5959 (1998).
2. Zlatev, I., Wang, L. & Steinhardt, P. J. *Phys. Rev. Lett.* **82**, 896–899 (1999).
3. Peebles, P. J. E. & Ratra, B. *Astrophys. J.* **325**, L17–L20 (1988).
4. Reiss, A. G. *et al.* *Astron. J.* **116**, 1009–1038 (1998).
5. Perlmutter, S. *et al.* <http://xxx.lanl.gov/abs/astro-ph/9812133>

## Immunology

# Accessory to murder

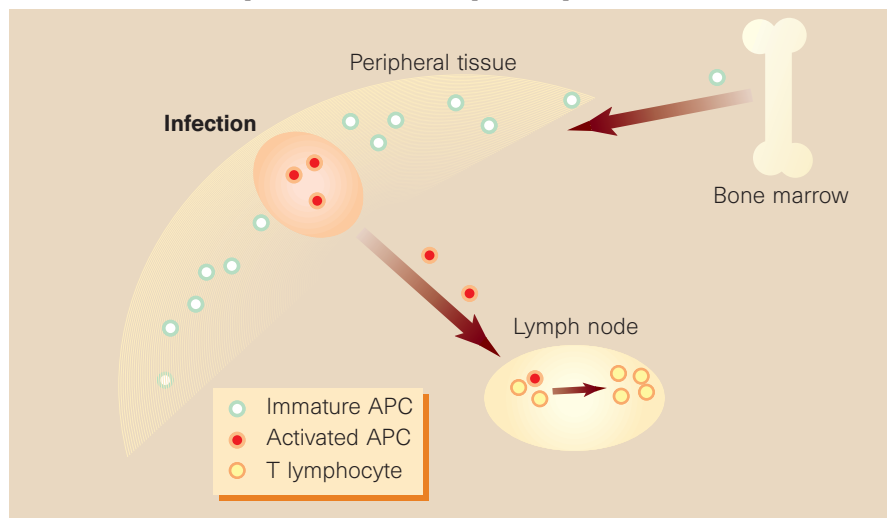
Ton N. M. Schumacher

Cytotoxic T cells rid the body of cells that have been invaded by viruses or bacteria. They do this by recognizing foreign peptide fragments associated with class I molecules of the major histocompatibility complex (MHC) in the membrane of the affected cells. The biggest challenge for the T cells lies not so much in killing the infected cells, as in tracking them down. The body contains  $10^{13}$  nucleated cells whereas, for a given viral antigen, there are perhaps a few hundred naive T cells — clearly, these few T cells could not scan all nucleated cells for signs of infection. In fact, naive T cells don't even try to enter peripheral tissues, and they are activated within the lymph nodes and spleen (immunologists consider all cells that are not derived from the bone marrow or lymphoid tissue to be peripheral). How, then, can a rare, virus-specific T cell in the

lymphoid compartment detect viral antigens produced in distant locations such as the skin? Sigal *et al.*<sup>1</sup> provide evidence (on page 77 of this issue) that bone-marrow-derived antigen-presenting cells (APCs) are the messengers that convey the news.

Over 40 years ago it was observed<sup>2</sup> that, during a local allergic reaction, initiation of the immune response requires the lymphatic vessels that drain the skin. Subsequent experiments pointed to migratory cells (that is, APCs) as the cells that transfer the antigen from the skin to the lymph node. So, for viruses that infect such migratory cells, the picture seems clear. But many viruses do not infect these migratory APCs. How does the T cell see these viruses?

As a rule, MHC-bound peptide fragments are derived from virus-encoded proteins produced inside the infected cell



**Figure 1** Detection of viral infection in peripheral tissues, based on the results of Sigal *et al.*<sup>1</sup>. Bone-marrow-derived antigen-presenting cells (APCs) are produced from stem cells in the bone marrow and, subsequently, migrate to peripheral tissues. Following infection, APCs in the infected tissues take up viral proteins or infected cells. After processing, viral antigens are displayed on the cell surface in association with major histocompatibility complex class I and class II molecules<sup>4,5</sup>. When virus-specific helper and killer T cells encounter these activated APCs in the draining lymph node, the result will be a virus-specific cellular immune response.

(endogenously). This seems appropriate, because it prevents inadvertent killing of innocent bystander cells that might have taken up viral or bacterial proteins. But, in landmark transplantation studies by Bevan<sup>3</sup>, cytotoxic T cells that recognize a foreign antigen in combination with a host MHC molecule were shown to be activated by foreign cells that lacked host MHC molecules. This phenomenon — historically known as cross-priming — provided the first, indirect, evidence for uptake and presentation of exogenous antigens, in the context of MHC class I, by host APCs.

In recent years, various *in vivo* models have been used to show the contribution of such exogenous presentation to cytotoxic T-cell responses against antigens expressed outside the lymphoid system (reviewed in refs 4, 5). These studies have provided solid evidence for the once heretical view that, in certain cell types, MHC class I molecules may sample antigens from the outside world. However, the requirements for T-cell activation in these model systems may differ considerably from those during a viral or bacterial infection<sup>6</sup>. So, the relevance of exogenous antigen presentation during a physiological encounter with a pathogen has long remained elusive.

Now, Sigal and colleagues<sup>1</sup> provide direct evidence that exogenous presentation is essential for induction of antiviral cytotoxic T-cell responses. They elegantly tackled the problem by making chimaeric mice in which all tissues except for those derived from the bone marrow contain the gene for the human poliovirus receptor. In these chimaeric mice, infection of cells with poliovirus (which does not normally infect mouse cells) can be tightly controlled. Using this strategy, the authors show that bone-marrow-derived APCs are essential for induction of an antiviral cytotoxic T-cell response. These APCs require a functional antigen-processing machinery, even when they cannot themselves be infected. The only straightforward explanation is that bone-marrow-derived APCs that are present in the infected peripheral tissues have taken up either viral antigens or infected cells, and displayed these in fragmented form for the scrutiny of naive cytotoxic T cells (Fig. 1). So, exogenous presentation seems to be crucial for induction of an immune response during viral infection of peripheral tissues.

Perhaps the prime reason we need this elaborate mechanism of T-cell activation is to facilitate the encounter between antigens that are produced in the periphery, and T cells located in the lymphoid compartment. But creation of a dedicated cell type to fulfil this task may bring one or two additional advantages to the host. Many viruses try to evade the host T-cell response by tinkering with the antigen-presentation machinery inside the infected cell<sup>7</sup>. Exogenous presen-

tation should ensure that viral escape mechanisms will not interfere with induction of a T-cell response, although these attempts will obviously still lessen the efficiency with which the resulting population of cytotoxic T cells will kill the infected cells.

Maybe more importantly, the APCs that gather and present foreign antigens have more than just a passive effect on T-cell activation. The APCs involved are probably dendritic cells and, over the past few years, a large body of work has indicated that dendritic cells act as cellular integrators. That is, by combining stimulatory and inhibitory signals delivered by B and T cells, cytokines, and viral and bacterial components, dendritic cells decide whether or not to mount a T-cell response<sup>8</sup>. This level of regulation may be essential for discrimination between self and non-self.

Sigal *et al.*<sup>1</sup> provide solid evidence that, by gathering and presenting antigens from virally infected cells, bone-marrow-derived APCs are an essential accessory to killer T cells. Although evidence has been obtained for the various modes by which exogenously

derived antigens are processed by APCs, our understanding of the molecular mechanisms that govern this route of antigen presentation is still limited<sup>4,5,9</sup>. It will be a challenge to work out how these professional messengers manage the remarkable feat of presenting, to the T cell, antigens that were never their own. □

Ton N. M. Schumacher is in the Division of Immunology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

e-mail: tschum@nki.nl

1. Sigal, L. J., Crotty, S., Andino, R. & Rock, K. L. *Nature* **398**, 77–80 (1999).
2. Fey, J. R. & Wenk, P. *Int. Arch. Allergy Immunol.* **11**, 81–94 (1957).
3. Bevan, M. J. *J. Immunol.* **117**, 2233–2238 (1976).
4. Rock, K. L. *Immunol. Today* **17**, 131–137 (1996).
5. Carbone, F. R., Kurts, C., Bennett, S. R., Miller, J. F. & Heath, W. R. *Immunol. Today* **19**, 368–373 (1998).
6. Bachmann, M. F., Zinkernagel, R. M. & Oxenius, A. *J. Immunol.* **161**, 5791–5794 (1998).
7. Ploegh, H. L. *Science* **280**, 248–253 (1998).
8. Lanzavecchia, A. *Nature* **393**, 413–414 (1998).
9. Jondal, M., Schirmeck, R. & Reimann, J. *Immunity* **5**, 295–302 (1996).

## Oceanography

# The mystery of the sapropels

Connie Sancetta

The Mediterranean Sea has an almost overwhelming amount of history, as anyone can attest who has tackled Fernand Braudel's great opus<sup>1</sup>. But even Braudel stopped short of describing the strange and wonderful geological history of the sea, which includes a sustained period of almost total desiccation during the latest Miocene, some 7–5 million years ago, and the puzzling occurrence of sapropels (Fig. 1). Sapropels are sediment layers with high concentrations of organic carbon, and for years there has been a simmering debate over the mechanism or mechanisms that could produce them. On page 57 of this issue<sup>2</sup>, Kemp *et al.* bring together a variety of observations from modern studies and sapropel microstructure to produce a feasible scheme of events which reconciles the apparent contradictions in the debate.

The facts of the case are simply stated. In the eastern Mediterranean, sediments of Plio-Pleistocene age (5 million to 10,000 years old) contain quasi-periodic laminated intervals whose lithology and geochemistry indicate that they accumulated under conditions of low — or even no — oxygen. Because the bottom waters are generally well oxygenated today, the puzzle has been: what conditions in the past reduced the oxidation state?

In principle, there are two ways to do this. First (option 1), sequester the deep water from atmospheric exchange for long enough that all of the dissolved oxygen is used up and

the deep water stagnates through the oxidation of accumulated organic matter. This means turning off the regional vertical circulation, effectively putting a lid on the system. In the salty Mediterranean, the most likely way to do this is by adding a surface layer of



Figure 1 A sapropel. These layers of sediment rich in organic carbon form in low-oxygen conditions. Kemp *et al.*<sup>2</sup> show that Mediterranean sapropels consist of alternating layers of two diatom assemblages. During periods when surface waters are stable, diatom mats form. Their subsequent rapid sinking deposits massive amounts of organic matter in the sediment, consuming the available oxygen.

A. E. S. KEMP

relatively fresh water, either from massive melting of northern ice sheets funnelled through the Black Sea<sup>3</sup> or from greatly increased run-off of the Nile River<sup>4</sup>.

Alternatively (option 2), there might be no difference in physical structure and circulation, but massive fluxes of organic matter to the sediment could overwhelm the amount of oxygen supplied by normal bottom-water circulation. In this case, the oxygen content of the water itself could be invariant, with oxygen consumption occurring within the surface sediment. This would imply much higher rates of primary production as a source of the organic matter<sup>5</sup>.

For years, there has been a tendency for those studying sapropels to assume that the processes are mutually exclusive: either stratification or production. Howell and Thunell<sup>6</sup> were the first to point out that this is a false dichotomy; it is an axiom of oceanography that high production results from stabilization of the upper water column, so the two processes are linked. Since their work, a more conciliatory tone has crept into publications, with authors at least paying lip service to the possibility of complementary, rather than opposing, conditions. It remains a minor theme, however, and one still reads papers referring exclusively to only one of the alternatives<sup>7,8</sup>.

Into this debate step Kemp *et al.*<sup>2</sup>, who weave together a detailed study of diatom-bearing sapropels using electron microscopy, and a range of observations from modern studies of plankton ecology. They show that the sapropels consist of regularly alternating pairs of laminae, one consisting of almost monospecific colonial diatoms and the other of a mixed diatom assemblage with other minor components. They explain this structure as resulting from a seasonal cycle in which the mixed-composition laminae represent normal winter–spring production. The monospecific laminae consist of several species with a distinctive ecology. These taxa harbour endosymbiotic nitrogen-fixing bacteria, so that the diatoms can continue growing when ambient waters are depleted of nutrients, as occurs during times of prolonged stratification. Moreover, these diatoms can form large mats, which aggregate at depth and can rise to the surface under calm (stratified) conditions.

Kemp *et al.* therefore propose that the mats formed during summer seasons of prolonged stratification (resulting from a density contrast produced by Nile run-off, option 1), and that rapid sinking of the mats at the beginning of autumn–winter wind mixing produced a massive load of organic material to the sediments that consumed the available oxygen (option 2). From mass-balance calculations, they argue that all of the organic carbon retained in the sapropels could have been supplied by the diatoms.

The authors have neatly reconciled the

supposedly conflicting processes of stratification and high production, producing a feasible conceptual model firmly grounded in modern observations. Although the model can apply only to diatomaceous sapropels, the authors correctly point out that the silica of these diatoms is highly soluble, and that sapropels lacking diatoms may once have possessed them. The mechanism might therefore apply to other carbon-rich laminated strata, such as the 120–100-million-year-old mid-Cretaceous black shales. These muds, containing up to 30% organic matter, are found in all the ocean basins, as well as in continental deposits from shallow interior seas. So diverse are their compositions and geological settings that it is difficult to explain their formation by any single mechanism. The processes proposed by Kemp *et al.*<sup>2</sup> now join the list of other concep-

tual models which, in essence, divide into vertical stratification and excess carbon flux — just like those suggested for their younger Mediterranean cousins. □

Connie Sancetta is in the Division of Ocean Sciences, Room 725, National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA.

e-mail: csancett@nsf.gov

1. Braudel, F. *The Mediterranean and the Mediterranean World in the Age of Philip II* (Harper & Row, New York, 1976).
2. Kemp, A. E. S., Pearce, R. B., Koizumi, I., Pike, J. & Rance, S. J. *Nature* **398**, 57–61 (1999).
3. Olausson, E. *Reps Swedish Deep-Sea Exped. VIII*, 336–391 (1961).
4. Rossignol-Strick, M., Nesteroff, W., Olive, P. & Vergnaud-Grazzini, C. *Nature* **295**, 105–110 (1982).
5. Calvert, S. *Oceanol. Acta* **6**, 255–267 (1983).
6. Howell, M. W. & Thunell, R. C. *Mar. Geol.* **103**, 461–471 (1992).
7. Béthoux, J.-P. & Pierre, C. *Mar. Geol.* **153**, 29–39 (1999).
8. Passier, H. F. *et al.* *Nature* **397**, 146–149 (1999).

#### Cell signalling

## DREAM on without calcium

Gail Mandel and Richard H. Goodman

On page 80 of this issue<sup>1</sup>, Carrión and co-workers describe an entirely new mechanism for the regulation of gene expression by an increased concentration of calcium ions in response to an incoming cellular signal. The authors have studied transcription from the gene encoding prodynorphin, an opioid peptide, and find that so long as the gene is tightly bound by a protein they call DREAM, transcription is repressed; but when calcium ions occupy their four tailor-made sites on DREAM, it floats off the DNA and allows the neuro-peptide to be expressed.

DREAM is a transcriptional repressor (its full name is 'downstream-regulatory-element antagonist modulator') that was identified through its ability to bind to the downstream-regulatory-element (DRE) sequence of the prodynorphin gene. Surprisingly, DREAM contains four sequence motifs known as 'EF-hands', which are often found in proteins that bind calcium ions. Typically, the binding of  $\text{Ca}^{2+}$  to EF-hands induces structural changes that alter the function of the protein. In the case of DREAM, binding of  $\text{Ca}^{2+}$  to its EF-hands leads to a reduced affinity for the target DRE sequence. Because DREAM is a repressor, the functional consequence is increased expression of the prodynorphin gene (Fig. 1, overleaf).

Direct binding to transcription factors is a departure from the previously characterized actions of  $\text{Ca}^{2+}$ . Calcium is known to modulate the activity of transcription factors by first activating cascades of protein kinases or phosphatases, which respectively phosphorylate and dephosphorylate a succession of intermediates. One kinase pathway involves CREB (for cyclic AMP-regulated enhancer

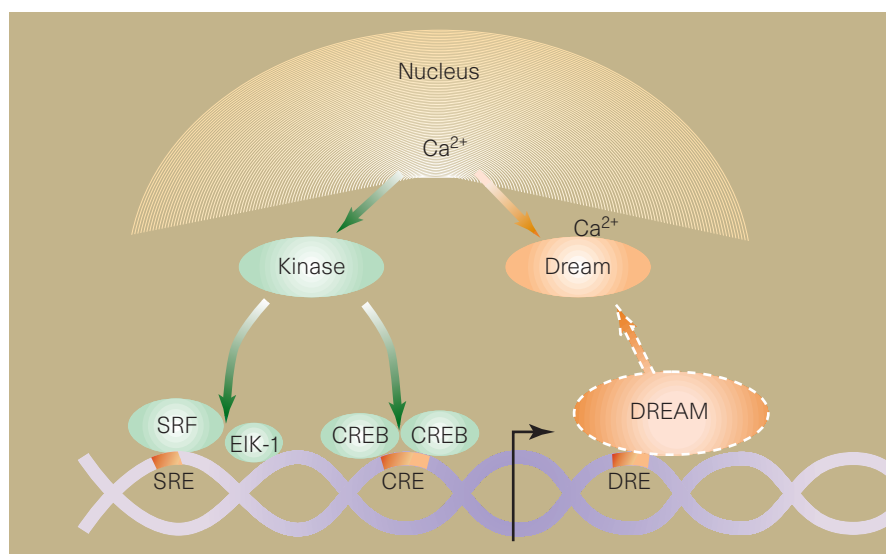
(CRE)-binding) protein, which was initially characterized as a mediator of extracellular signals using the cAMP/protein kinase A (PKA) pathway. A rise in intracellular calcium activates CREB by a different route, by triggering the sequential activation of another set of kinases,  $\text{Ca}^{2+}$ /calmodulin (CaM) kinase and CaM kinase IV (ref. 2).

CaM kinase IV phosphorylates the same residue in CREB as PKA does, allowing the co-activator CREB-binding protein (CBP), or its homologue p300, to bind. CBP is itself activated by the CaM kinase IV pathway, which may augment the  $\text{Ca}^{2+}$  signal<sup>3</sup>. Conversely,  $\text{Ca}^{2+}$  regulation in lymphocytes of the protein complex known as NF-AT involves the phosphatase calcineurin, which dephosphorylates NF-AT and causes shuttling of the transcriptional activator into the nucleus<sup>4</sup>. Modulation of a transcriptional repressor by direct binding to  $\text{Ca}^{2+}$  is the latest twist in the tale of how calcium signals are sensed in the nucleus.

The presence of EF-hands in DREAM raises questions concerning its mechanism of action. For example, is  $\text{Ca}^{2+}$  the only mediator of DREAM regulation, or could  $\text{Mg}^{2+}$  fulfil the same function? The second EF-hand in DREAM contains an aspartate instead of glutamate in the Z position (one of six residues that contribute an oxygen atom to the metal-ion coordination). Studies of the smooth-muscle myosin regulatory light chain indicate that this substitution might expand the binding affinity from  $\text{Ca}^{2+}$  alone to  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  together<sup>5</sup>.

Magnesium is not normally considered as a regulator, but measurements using the  $\text{Mg}^{2+}$ -sensitive dye Magfura-2 have detected large changes in free  $\text{Mg}^{2+}$  concentrations in





**Figure 1** The transcriptional repressor DREAM is regulated by direct binding to calcium. In the basal state, DREAM is bound to DRE located downstream from the transcriptional start site (black arrow). When the intracellular  $\text{Ca}^{2+}$  concentration rises, DREAM binds to  $\text{Ca}^{2+}$  and comes off the DRE, permitting a higher level of promoter activity. In contrast to DREAM, the transcriptional activators CREB and SRF/Elk-1 sense  $\text{Ca}^{2+}$  through the action of intermediary kinases, and remain bound to the Ca/CRE and SRE sites, respectively, in the presence of  $\text{Ca}^{2+}$ .

cortical neurons after treatment with the neurotransmitter glutamate<sup>6</sup>. The magnitude of these changes might be sufficient to affect DREAM structure and DNA binding. For other proteins, such as the neuronal calcium-binding proteins VILIP and NCS-1, binding of  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  has very different effects: NCS-1 structure is altered by binding of either  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$ , whereas binding of  $\text{Mg}^{2+}$  to VILIP does not cause any structural change<sup>7</sup>.

Indeed, the finding that DREAM binds  $\text{Ca}^{2+}$  at all comes as a surprise, because the DRE was first characterized by its ability to mediate PKA-dependent de-repression of prodynorphin gene expression<sup>8</sup>. Carrión *et al.* had previously shown that PKA decreases binding of DREAM to the DRE<sup>8</sup>. In their latest report, forskolin had no effect on DREAM function; also, the DREAM complementary DNA sequence does not predict a consensus PKA-recognition site. It is possible that PKA might phosphorylate DREAM through another kinase pathway. Alternatively, the 110-kilodalton protein regulated by PKA might contain DREAM as a subunit; or the PKA effects they observed in their earlier study could be a result of an alteration in the function of  $\text{Ca}^{2+}$  channels.

Now that the DREAM complementary DNA has been isolated, it should be possible to generate antibodies to determine the properties of the endogenous DREAM protein in neuronal cells. This is important because regulation by DREAM is not likely to be restricted to the prodynorphin gene. The DRE sequence is found in at least one other gene regulated by  $\text{Ca}^{2+}$ , the immediate-early gene *c-fos*. Carrión *et al.*<sup>1</sup> have shown that disruption of the EF-hands in DREAM, or mutation

of the DRE, blocks the de-repression of *c-fos* reporter genes that normally occurs in response to treatment with caffeine, which raises the intracellular  $\text{Ca}^{2+}$  concentration.

However, the relative contribution of the DRE in the context of other regulatory elements is not assessed in the new study<sup>1</sup>. This is an important point, because two other well characterized calcium-responsive elements, the CARE/CRE and the serum-response element (SRE), contribute to the regulation of the *c-fos* gene<sup>9</sup>. Thus, the physiological function of DREAM is still uncertain. Perhaps J. A. Hadfield<sup>10</sup> had it right when he answered his own question: "Would we say that the function of the dream is to express something or to hide something? It is both at once." But he was probably not thinking of prodynorphin. □

Gail Mandel is at the Howard Hughes Medical Institute, Department of Neurobiology and Behavior, State University of New York, Stony Brook, New York 11794-5230, USA.  
e-mail: gmandel@brain.neurobio.sunysb.edu  
Richard H. Goodman is at the Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA.

1. Carrión, A. M., Link, W. A., Ledo, F., Mellstrom, B. & Naranjo, J. R. *Nature* **398**, 80–84 (1999).
2. Tokumitsu, H. *et al.* *J. Biol. Chem.* **270**, 19320–19324 (1995).
3. Chawla, S. *et al.* *Science* **281**, 1505–1509 (1998).
4. Crabtree, G. R. & Clipstone, N. A. *Annu. Rev. Biochem.* **63**, 1045–1083 (1994).
5. DaSilva, A. C. R., Kendrick-Jones, J. & Reinach, F. C. *J. Biol. Chem.* **270**, 6773–6778 (1995).
6. Brocard, J. B., Rajdev, S. & Reynolds, I. J. *Neuron* **11**, 751–757 (1993).
7. Cox, J. A. *et al.* *J. Biol. Chem.* **269**, 32807–32813 (1994).
8. Carrión, A. M., Mellstrom, B. & Naranjo, J. R. *Mol. Cell. Biol.* **18**, 6921–6929 (1998).
9. Finkbeiner, S. & Greenberg, M. E. *BioEssays* **19**, 657–660 (1997).
10. Hadfield, J. A. in *Dreams and Nightmares* 139 (Penguin, Harmondsworth, UK, 1954).

Daedalus

## Bugging the brain

Last week Daedalus outlined his new 'microdermic needle' injection technology. A giant multilayered carbon nanotube, a micrometre across and coated with slippery graphite fluoride, is drawn into the body by ultrasonic pulses sent down it. It noses its way harmlessly between the cells, even those of bone, and is steered under imaging control to the injection site.

Daedalus now muses that, thanks to its insulating coating of graphite fluoride, his new needle is an ideal biological electrode. It could sense an electrical potential at its tip, and relay electrical signals back down again.

The obvious target organ is the brain. Unlike electroconvulsive therapy, microdermic therapy would allow precise intervention. Far too fine to hurt or even be noticed, a microdermic needle could be inserted into the skull of the conscious patient, and steered around his brain. Like a precision electroencephalograph, it would detect the electrical activity of the cells it encountered. Pulses could then be sent down it to reinforce or modify their action. Sudden memories, ideas, or motives might grip the patient, revealing their exact site in the brain. His mentality could be mapped in detail and its troubles precisely located.

The same needle could then be used for therapy. A drug — say a neurotransmitter or an antipsychotic — could be passed down it to normalize the local brain cells. Alternatively, a pattern of electrical pulses might be devised for the same purpose. Inert, painless, undetectable, the needle could be left in place indefinitely, for future (or even maintained 'depot') corrective action. Manias, obsessions and delusions could be erased, tics and seizures switched off. Psychiatry would become an exact science at last.

The technology might even be developed into quite a general man-machine interface. Inserted into a nerve, a microdermic needle could read the traffic in that nerve, and launch impulses into it. A computer attached to the needle could learn to command that nerve. The first applications would be prosthetic — restoring sensation and action to victims of paralysis, or sensory and functional disorders. But a mass of needles densely distributed through the nervous system, all connected to a true back-pack 'personal computer', would create the very first cyborg robot: a computer seeing through human eyes and controlling a human body.

David Jones

# Designing tie knots by random walks

The simplest of conventional tie knots, the four-in-hand, has its origins in late-nineteenth-century England. The Duke of Windsor, as King Edward VIII became after abdicating in 1936, is credited with introducing what is now known as the Windsor knot, from which its smaller derivative, the half-Windsor, evolved. In 1989, the Pratt knot, the first new knot to appear in fifty years, was revealed on the front page of *The New York Times*.

Rather than wait another half-century for the next sartorial advance, we have taken a more formal approach. We have developed a mathematical model of tie knots, and provide a map between tie knots and persistent random walks on a triangular lattice. We classify knots according to their size and shape, and quantify the number of knots in each class. The optimal knot in a class is selected by the proposed aesthetic conditions of symmetry and balance. Of the 85 knots that can be tied with a conventional tie, we recover the four knots that are in widespread use and introduce six new aesthetically pleasing knots.

A tie knot is started by bringing the wide (active) end to the left and either over or under the narrow (passive) end, dividing the space into right (R), centre (C) and left (L) regions (Fig. 1a). The knot is continued by subsequent half-turns, or moves, of the active end from one region to another (Fig. 1b) such that its direction alternates between out of the shirt ( $\odot$ ) and into the shirt ( $\otimes$ ). To complete a knot, the active end must be wrapped from the right (or left) over the front to the left (or right), underneath to the centre and finally through (denoted T but not considered a move) the front loop just made.

Elements of the move set  $\{R_{\odot}, R_{\otimes}, C_{\odot}, C_{\otimes}, L_{\odot}, L_{\otimes}\}$  designate the moves necessary to place the active end into the corresponding region and direction. We can then define a tie knot as a sequence of moves initiated by  $L_{\odot}$  or  $L_{\otimes}$  and terminating with the subsequence  $R_{\odot}L_{\otimes}C_{\otimes}T$  or  $L_{\otimes}R_{\odot}C_{\odot}T$ . The sequence is constrained such that no two consecutive moves indicate the same region or direction.

We represent knot sequences as random walks on a triangular lattice (Fig. 1c). The axes  $r$ ,  $c$  and  $l$  correspond to the three move regions R, C and L, and the unit vectors  $\hat{r}$ ,  $\hat{c}$  and  $\hat{l}$  represent the corresponding moves; we omit the directional notation  $\odot, \otimes$  and the terminal action T. Because all knot sequences end with  $C_{\odot}$  and alternate between  $\odot$  and  $\otimes$ , all knots of odd numbers of moves begin with  $L_{\odot}$ , whereas those of even numbers of moves begin with  $L_{\otimes}$ .

Our simplified random-walk notation is therefore unique.

The size of a knot, and the primary parameter by which we classify it, is the number of moves in the knot sequence, denoted by the half-winding number  $h$ . The initial and terminal sequences dictate that the smallest knot is given by the sequence  $L_{\odot}R_{\otimes}C_{\odot}T$ , with  $h=3$ . Practical considerations (namely the finite length of the tie), as well as aesthetic ones, suggest an upper bound on knot size, so we limit our exact results to  $h \leq 9$ .

The number of knots as a function of size,  $K(h)$ , corresponds to the number of walks of length  $h$  beginning with  $\hat{l}$  and ending with  $\hat{r}\hat{l}\hat{c}$  or  $\hat{l}\hat{r}\hat{c}$ . It may be written

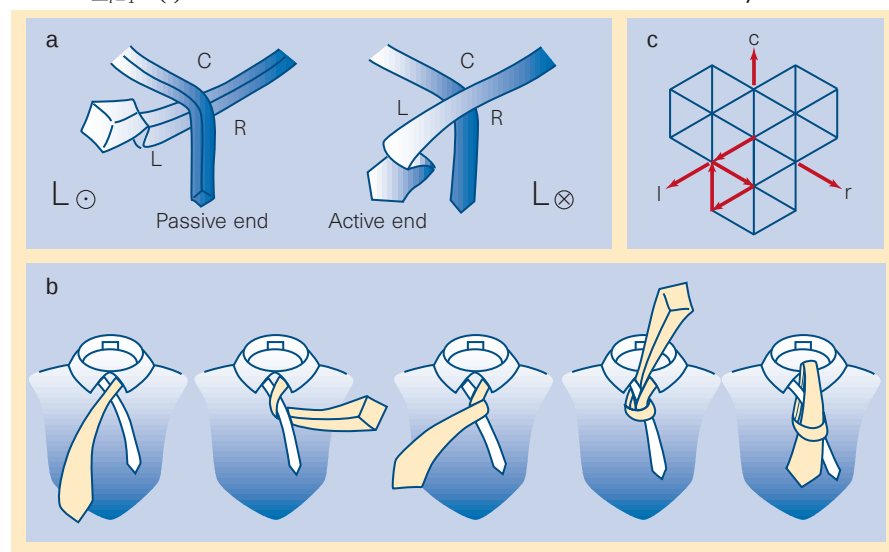
$$K(h) = (1/3)(2^{h-2} - (-1)^{h-2})$$

where  $K(1)=0$ , and the total number of knots is  $\sum_{i=1}^9 K(i) = 85$ .

The shape of a knot depends on the number of right, centre and left moves in the tie sequence. Because symmetry dictates that there be an equal number of right and left moves (see below), the shape of a knot is characterized by the number of centre moves  $\gamma$ . We use it to classify knots of equal size  $h$ ; knots with identical  $h$  and  $\gamma$  belong to the same class. A large centre fraction  $\gamma/h$  indicates a broad knot (such as the Windsor) and a small centre fraction suggests a narrow one (such as the four-in-hand), but not all centre fractions allow aesthetic knots. We therefore limit our attention to  $1/4 \leq \gamma/h \leq 1/2$ .

The number of knots in a class,  $K(h, \gamma)$ , is equivalent to the number of walks of length  $h$  that satisfy the boundary conditions and contain  $\gamma$  steps  $\hat{c}$ ; it appears as

$$K(h, \gamma) = 2^{\gamma-1} \binom{h-\gamma-2}{\gamma-1}$$



**Figure 1** All diagrams are drawn in the frame of reference of the mirror image of the actual tie. a, The two ways of beginning a knot,  $L_{\odot}$  and  $L_{\otimes}$ . For knots beginning with  $L_{\odot}$ , the tie must begin inside-out. b, The four-in-hand, denoted by the sequence  $L_{\otimes}R_{\odot}L_{\odot}C_{\odot}T$ . c, A knot may be represented by a persistent random walk on a triangular lattice. The example shown is the four-in-hand, indicated by the walk  $\hat{l}\hat{r}\hat{l}\hat{c}$ .

| $h$ | $\gamma$ | $\gamma/h$ | $K(h, \gamma)$ | $s$ | $b$ | Name         | Sequence                                                                                       |
|-----|----------|------------|----------------|-----|-----|--------------|------------------------------------------------------------------------------------------------|
| 3   | 1        | 0.33       | 1              | 0   | 0   |              | $L_{\odot}R_{\otimes}C_{\odot}T$                                                               |
| 4   | 1        | 0.25       | 1              | -1  | 1   | Four-in-hand | $L_{\otimes}R_{\odot}L_{\odot}C_{\odot}T$                                                      |
| 5   | 2        | 0.40       | 2              | -1  | 0   | Pratt knot   | $L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$                                           |
| 6   | 2        | 0.33       | 4              | 0   | 0   | Half-Windsor | $L_{\otimes}R_{\odot}C_{\otimes}L_{\otimes}R_{\odot}C_{\odot}T$                                |
| 7   | 2        | 0.29       | 6              | -1  | 1   |              | $L_{\otimes}R_{\odot}L_{\otimes}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$                     |
| 7   | 3        | 0.43       | 4              | 0   | 1   |              | $L_{\odot}C_{\otimes}R_{\odot}C_{\otimes}L_{\otimes}R_{\odot}C_{\odot}T$                       |
| 8   | 2        | 0.25       | 8              | 0   | 2   |              | $L_{\otimes}R_{\odot}L_{\otimes}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$                     |
| 8   | 3        | 0.38       | 12             | -1  | 0   | Windsor      | $L_{\odot}C_{\otimes}R_{\odot}L_{\otimes}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$            |
| 9   | 3        | 0.33       | 24             | 0   | 0   |              | $L_{\otimes}R_{\odot}C_{\otimes}L_{\otimes}R_{\odot}C_{\otimes}L_{\otimes}R_{\odot}C_{\odot}T$ |
| 9   | 4        | 0.44       | 8              | -1  | 2   |              | $L_{\odot}C_{\otimes}R_{\odot}C_{\otimes}L_{\otimes}C_{\otimes}R_{\odot}L_{\otimes}C_{\odot}T$ |

Knots are characterized by half-winding number  $h$ , centre number  $\gamma$ , centre fraction  $\gamma/h$ , knots per class  $K(h, \gamma)$ , symmetry  $s$ , balance  $b$ , name and sequence.

The symmetry of a knot, which is our first aesthetic constraint, is determined by the number of moves to the right minus the number of moves to the left,

$$s = \sum_{i=1}^h x_i$$

where  $x_i = 1$  if the  $i$ th step is  $\hat{R}$ ,  $-1$  if the  $i$ th step is  $\hat{L}$  and  $0$  otherwise. Because asymmetric knots disrupt human bilateral symmetry, we consider the most symmetric knots from each class, that is, the ones that minimize  $s$ .

Whereas the centre number  $\gamma$  and the symmetry  $s$  specify the move composition of a knot, balance relates to the distribution of these moves; it corresponds to the extent to which the moves are mixed. A balanced knot is tightly bound and keeps its shape. We use this as our second aesthetic constraint. The balance  $b$  may be expressed as

$$b = (1/2) \sum_{i=2}^{h-1} |\omega_i - \omega_{i-1}|$$

and the winding direction  $\omega_i(\sigma_i, \sigma_{i+1}) = 1$ , where  $\sigma_i$  represents the  $i$ th step of the walk, if the transition from  $\sigma_i$  to  $\sigma_{i+1}$  is clockwise, say, and  $-1$  otherwise. Of those knots that are optimally symmetric, we desire that knot which minimizes  $b$ .

The ten canonical knot classes  $\{h, \gamma\}$  and the corresponding most aesthetic knots are listed in Table 1. The four named knots are the only ones, to our knowledge, to have received widespread attention, either published or through tradition. Here we introduce some unnamed knots.

The first four columns of Table 1 describe the knot class  $\{h, \gamma\}$ , whereas the remainder relate to the corresponding most aesthetic knot. The centre fraction  $\gamma/h$  provides a guide to the shape of a knot, with higher fractions corresponding to broader knots; along with the size  $h$ , it should be used in selecting a knot.

Some readers may notice the use of knots whose sequences are equivalent to those shown in Table 1 apart from transpositions of  $\hat{R}$ ,  $\hat{L}$  groups, such as the use of  $L_{\odot} R_{\odot} C_{\odot} R_{\odot} L_{\odot} C_{\odot} T$  in place of the half-Windsor (T. P. Harte and L. S. G. E. Howard, personal communication); some will argue that this is the half-Windsor. Such ambiguity follows from the variable width of conventional ties (the earliest ties were uniformly wide). This makes some transpositions arguably favourable, namely the last  $\hat{R}$ ,  $\hat{L}$  group in the knots  $\{5, 2\}$ ,  $\{6, 2\}$ ,  $\{7, 2\}$ ,  $\{8, 3\}$  and  $\{9, 3\}$  in Table 1. We do not attempt to distinguish between these knots and their counterparts; this much we leave to the sartorial discretion of the reader.

**Thomas M. Fink, Yong Mao**

Cavendish Laboratory,  
Cambridge CB3 0HE, UK  
e-mail: tmf20@cus.cam.ac.uk

## Pasture damage by an Amazonian earthworm

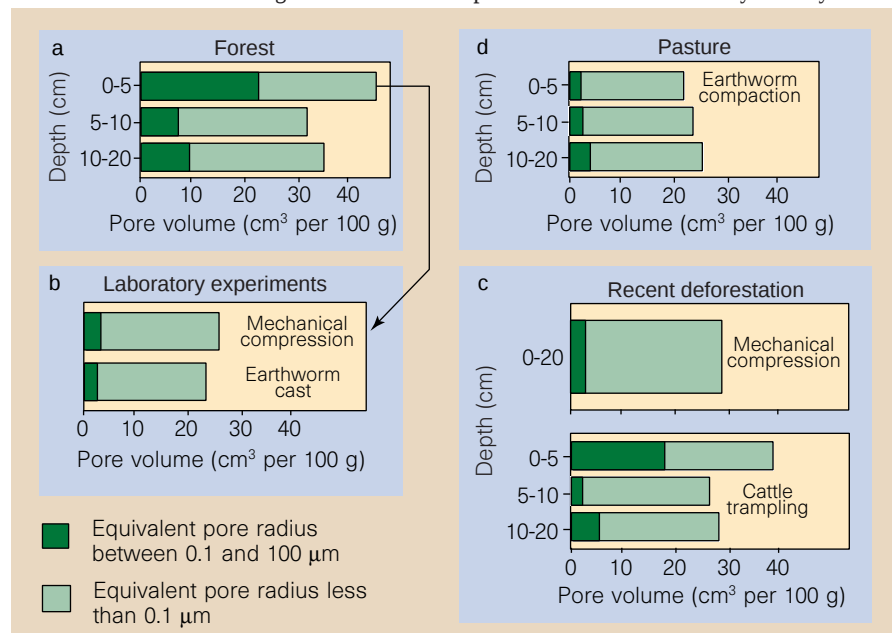
Almost all cultivated soils undergo some reduction in the porosity of the surface layers, and nowhere is this more evident than in tropical rainforests that have been converted to pastures. Following deforestation in an area of Costa Rica, soil bulk density has been shown to increase rapidly after conversion to pasture, leading to poor drainage and a reduced rate of gaseous diffusion<sup>1</sup>. These factors limit methane consumption and promote the anaerobic production of methane. A similar effect on methane flux has been found in upland soils in the Brazilian Amazonian basin after conversion from forest to pasture<sup>2,3</sup>. Increases in atmospheric methane are therefore not limited to emissions from flooded soils<sup>4</sup>, as forest-to-pasture conversion promotes the anaerobic mineralization of organic matter by changing the physical properties of soil.

We now demonstrate the importance of this process in pasture degradation in central Amazonia, close to Manaus in northern Brazil (Fig. 1a). We show that, in addition to the substantial compacting effects of heavy machinery<sup>5</sup> and cattle trampling<sup>6</sup>, another more insidious agent — the soil

macrofauna — can have profound and lasting effects<sup>7,8</sup> on the porosity of pasture soils.

We wetted a natural forest soil (xanthic acruox, USDA, 1996) to a water potential of  $-10$  kPa and compressed it at a pressure of  $10^3$  kPa, measured using an oedometer. Macroporosity (in the range  $0.1$  to  $100 \mu\text{m}$ ) fell from  $21.7$  to  $3 \text{ cm}^3$  per  $100 \text{ g}$ , indicating that forest surface soils (at a depth of  $0$ – $5 \text{ cm}$ ) are extremely sensitive to compaction (Fig. 1b). Passing the same soil through the gut of the earthworm *Pontoscolex corethrurus*, an aggressive exotic colonist that invades many tropical pastures, reduced macroporosity even more to  $1.6 \text{ cm}^3$  per  $100 \text{ g}$ . We believe that this change is brought about by the intense mixing and near-complete dispersion of soil particles in the moist environment of the earthworm gut (water content:  $0.85 \text{ g}$  per  $\text{g}$ ).

During conversion from forest to pasture, two separate mechanisms act to compact the soil. First, the effects of heavy machinery<sup>5</sup> and trampling by cattle<sup>6</sup> occur in specific locations and result from the techniques used for deforestation and pasture management. They lead to the mixing of plant debris and clay in the upper  $5 \text{ cm}$  and to severe compaction in the layer  $5$ – $10 \text{ cm}$  deep (Fig. 1c). Second, the effects of reduced abundance and diversity of macrofaunal communities in the newly created pastures are linked to ecosystem dynamics



**Figure 1** Effect of mechanical and biological action on soil pore distribution in central Amazonia. Data are deduced from mercury porosimetry analysis<sup>10</sup>. The pore size distribution shows a bimodal pattern, indicating that there are two types: the largest pores, which have an equivalent pore radius (EPR) between  $0.1$  and  $100 \mu\text{m}$ , are biological in origin or are fine fissures, and are essential for gas exchange and the infiltration and retention of free water; the smaller pores, with an EPR of less than  $0.1 \mu\text{m}$ , are found in compact clumps of kaolinite. **a**, In primary forest, biodiverse soil macrofauna and roots regulate soil pore volume. **b**, Laboratory experiments in which the forest surface horizon is compacted by mechanical compression ( $10^3$  kPa) of soil with a water potential of  $-10$  kPa, or by mechanical working and dispersion in the earthworm gut. **c**, Effects of recent deforestation due to heavy machinery and trampling by cattle (after manual clearance). **d**, Ungrazed, manually deforested pasture shows accumulation of compact surface casts in the layer  $0$ – $5 \text{ cm}$  deep.



and so are widespread and long-lasting. While the forest remains undisturbed, a highly diverse soil macrofauna maintains an adequate soil pore volume by balancing the effects of compacting and decompacting agents<sup>7</sup> (Fig. 1a). However, deforestation and the establishment of exotic grasses lead to a dramatic fall in biodiversity: 68% of the soil macrofaunal taxa of the original forest disappear and are replaced by large populations of the earthworm *P. corethrurus* (to a maximum of 400 per m<sup>2</sup>).

These earthworm populations can make up nearly 90% of soil invertebrate biomass, and produce more than 100 tonnes per hectare per year of casts, dramatically decreasing macroporosity to 2.7 cm<sup>3</sup> per 100 g, a level as low as that found after the compaction of moist soils by heavy machinery. During the rainy season (November to June), intense casting by *P. corethrurus* at the surface of the saturated soil leads to the formation of a large continuous grey mass of casts 5 cm thick with a water content as high as 0.65 g per g. While the soil remains moist, this muddy layer has a continuous water phase that impedes gas exchange between the soil and the atmosphere. The resulting anaerobic soil conditions lead to the increased production of methane and limit its consumption. At the end of the rainy season, desiccation cracks form in the drying soil and the planted pastures wilt.

The best way to restore soils with this type of degraded structure is to maintain or establish enough woody plants in the landscape<sup>9</sup> to support an adequately diverse macro-invertebrate fauna.

**Armand Chauvel\*, Michel Grimaldi†, Eleusa Barros‡, Eric Blanchart§, Thierry Desjardins\*, Max Sarrazin‡, Patrick Lavelle\***

\*Laboratoire d'Ecologie des Sols Tropicaux, ORSTOM, 32 avenue Henri Varagnat, 93143 Bondy Cedex, France

e-mail: chauvel@bondy.orstom.fr

†ORSTOM/INRA, Science du Sol, 65 rue de St Briec, 35000 Rennes, France

‡INPA/ORSTOM—Ecologia, CP 478, 69011-970, Manaus AM, Brazil

§ORSTOM Martinique, BP 8002, 97259 Fort de France, France

- Keller, M., Veldkamp, E., Weltz, A. M. & Reiners W. A. *Nature* **365**, 244–246 (1993).
- Steudler, P. A. et al. *J. Geophys. Res.* **101**, 18547–18554 (1996).
- Luizao, F., Matson, P., Luizao, R. & Vitousek, P. *Global Biogeochem. Cycles* **3**, 281–285 (1989).
- Seiler, W. & Conrad, R. in *The Geophysiology of Amazonia. Vegetation and Climate Interactions* (ed. Dickinson, R. E.) 133–162 (Wiley, New York, 1987).
- Chauvel, A., Grimaldi, M. & Tessier, D. *Forest Ecol. Mgmt* **38**, 259–274 (1991).
- Grimaldi, M. et al. *Cahiers Agric.* **2**, 36–47 (1993).
- Blanchart, E., Lavelle, P., Braudeau, E., Le Bissonais, Y. & Valentin, C. *Soil Biol. Biochem.* **29**, 431–439 (1997).
- Alegre, J. C., Pashanasi, B. & Lavelle, P. *Soil Sci. Soc. Am. J.* **60**, 1522–1529 (1996).
- Ramachandran Nair, P. K. in *An Introduction to Agroforestry* 261–343 (Kluwer, 1993).
- Lawrence, G. P. *J. Soil Sci.* **28**, 527–540 (1977).

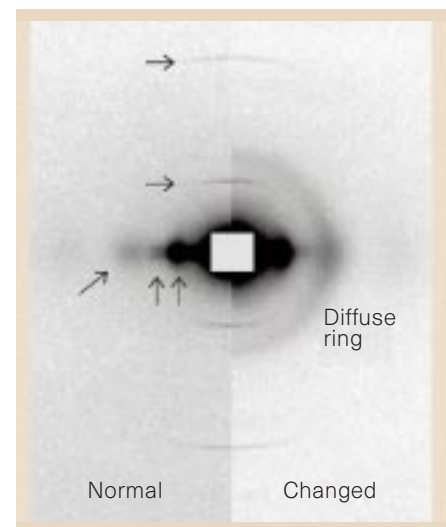
## Using hair to screen for breast cancer

We have studied hair using fibre X-ray diffraction studies with synchrotron radiation and find that hair from breast-cancer patients has a different intermolecular structure to hair from healthy subjects. These changes are seen in all samples of scalp and pubic hair taken from women diagnosed with breast cancer. All the hair samples from women who tested positive for a mutation of the *BRCA1* gene, which is associated with a higher risk of breast cancer<sup>1</sup>, also show these changes. Because our results are so consistent, we propose that such hair analyses may be used as a simple, non-invasive screening method for breast cancer.

We previously examined the intermolecular structure of normal human hair by using synchrotron X-ray scattering on multiple fibres<sup>2,3</sup>. The change in the hair from breast-cancer patients can be seen in the X-ray diffraction pattern as one or more rings of intensity that superimpose on the pattern for normal hair (Fig. 1). We believe that this pattern arises from a variation in the structure of the cell membrane as the hair is formed in the follicle.

We performed four studies on hair from controls and from patients with breast carcinoma. The first double-blind study (samples provided by A. Howell, Christie CRC Research Centre, Manchester) was carried out on beamline 15A at the Photon Factory (Japan). After correlating the diffraction patterns with patient data, we considered that the changes might indicate a propensity to malignancy and be related to breast cancer *per se*. To test this idea, we did another double-blind study on a larger sample of hair from the same source. This time, the X-ray measurements were made on the Australian National Beamline Facility and later confirmed on the BioCAT beamline at the Advanced Photon Source (Chicago).

The initial results showed some inconsistencies, which were eliminated when we excluded hair samples that had been 'permed' less than three months before collection. To avoid problems caused by hair treatment, we used samples of pubic hair collected at the Oncology Unit, St George



**Figure 1** X-ray scattering patterns. Left, hair from a healthy 21-year-old female; right, hair from a 48-year-old breast-cancer patient. Horizontal and vertical arrows, parts of the pattern that are unchanged and arise from ordering and packing in the hair fibril. Diagonal arrow, a region of intensity arising from the membrane bonding the fibrils together. In the changed pattern, a diffuse ring resulting from random orientation superimposes on this region.

Hospital (Sydney, Australia).

All hair samples (23 out of 23) taken from breast-cancer patients exhibited the characteristic change in their X-ray scattering patterns (Table 1). Of the samples taken from patients not suspected of having breast cancer, the scattering patterns of 86% (24 out of 28) were normal. In addition, of those not yet diagnosed with breast cancer but suspected of being at risk (because of familial history and the presence of a pathological mutation of *BRCA1*), 60% (3 out of 5) showed the full change associated with breast carcinoma; the remainder showed a partial change.

The ring that characterizes the hair from breast-cancer patients corresponds to a molecular spacing of  $4.44 \pm 0.06$  nm, which places it directly in the position of the equatorial arc representing the plasma membrane in the normal hair pattern. The ring probably arises from randomly orientated lipid bilayers, either in the plasma membrane or in membranous inclusions in the hair cells, but this needs further study.

We correctly identified all the samples from women with breast cancer, but further

| Breast cancer diagnosed? | Samples | Samples with changed structure | Familial history of cancer (pathologic <i>BRCA1</i> mutation) | Hair origin |
|--------------------------|---------|--------------------------------|---------------------------------------------------------------|-------------|
| Yes                      | 15      | 15                             | Yes (positive)                                                | Scalp       |
| Yes                      | 8       | 8                              | None                                                          | Pubic       |
| No                       | 3       | 3                              | Yes (positive)                                                | Scalp       |
| No                       | 2       | Partial 2                      | Yes (positive)                                                | Scalp       |
| No                       | 8       | 1                              | Yes (negative)                                                | Scalp       |
| No                       | 16      | 3                              | None                                                          | Scalp       |
| No                       | 4       | 0                              | None                                                          | Pubic       |

investigation into the sensitivity and specificity of this test is necessary. It is possible that this may lead to a simple and reliable screening method for breast cancer using a single pubic hair.

**Veronica James\***, **John Kearsley†**,  
**Tom Irving‡**, **Yoshiyuki Amemiya§**,  
**David Cookson¶**

\*Rex Vowels Low Angle Diffraction Laboratory,  
University of New South Wales,  
Sydney, New South Wales 2052, Australia  
e-mail: vjs@bigpond.com

†Cancer Care Centre, St George Hospital,  
Kogarah 2217, Australia

‡BioCAT, BCPS Department,  
Illinois Institute of Technology,  
Chicago, Illinois 60616, USA

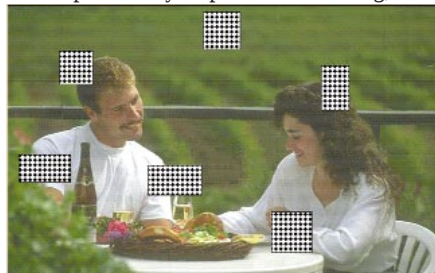
§Engineering Research Unit, School of Engineering,  
University of Tokyo, 2-11-16 Yayoi,  
Bunkyo-ku, Tokyo 113, Japan

¶Australian Nuclear Science and Technology  
Organisation, Private Mail Bag 1,  
Menai, New South Wales 2234, Australia

1. Struwing, J. P. et al. *N. Engl. J. Med.* **336**, 1401–1408 (1997).
2. Wilk, K. E., James, V. J. & Amemiya, Y. *Biochim. Biophys. Acta* **1245**, 392–396 (1995).
3. James, V. J. & Amemiya, Y. *Textile Res. J.* **68**, 167–170 (1998).

## Change-blindness as a result of 'mudsplashes'

Change-blindness<sup>1,2</sup> occurs when large changes are missed under natural viewing conditions because they occur simultaneously with a brief visual disruption, perhaps caused by an eye movement<sup>3,4</sup>, a flicker<sup>5</sup>, a blink<sup>6</sup>, or a camera cut in a film sequence<sup>7</sup>. We have found that this can occur even when the disruption does not cover or obscure the changes. When a few small, high-contrast shapes are briefly spattered over a picture, like mudsplashes on a car windscreen, large changes can be made simultaneously in the scene without being noticed. This phenomenon is potentially important in driving, sur-



**Figure 1** Mudsplashes consisted of six small black-and-white textured rectangles or ovals, dispersed over the picture so as not to cover the location of the change. The change in this example consisted of the bar in the background moving down and back up. This was judged to be a 'marginal interest' change by a panel of six judges in a previous pilot experiment. Movies of the effects, and further details can be found in Supplementary Information.

veillance or navigation, as dangerous events occurring in full view can go unnoticed if they coincide with even very small, apparently innocuous, disturbances. It is also important for understanding how the brain represents the world.

We used 48 pairs of pictures, consisting of an original and a modified picture (each displayed for 3 s), each pair presented cyclically with an 80-ms duration 'mudsplash' (Fig. 1) superimposed at the moment of the change. There was no disruption in visual continuity at the moment of the change. Ten observers were asked to press a button as soon as they identified the change, which could have been a large object or region of the picture shifting in location, changing colour, or appearing or disappearing. The change could be either a 'central interest' or a 'marginal interest' element (Fig. 1). Both types of change were equalized for size and salience and were visible under normal conditions without the mudsplash.

Central-interest changes were usually detected as soon as they occurred, whereas marginal-interest changes were seen only on their second or later occurrences. In 13–30% of the cases, marginal-interest changes, although in full view, were not detected at all during the 40-s viewing period.

Part of the explanation for these effects is that attention-grabbing luminance transitions, caused all over the visual field by the brief visual disruptions, prevent attention being focused on the location of the change. But the question remains as to why, once the extraneous transients have subsided, comparison between the current and previous views of the scene is still impossible.

Because the mudsplashes provoke only a minor disturbance, and because they do not cover the location of the change, the change-blindness they cause cannot be due to masking, or to erasure or resetting of the information contained in an internal representation of the visual world. Rather, it seems that change-blindness occurs because the internal representation of the visual world is rather sparse and essentially contains only central-interest information. A second experiment confirmed this.

Instead of mudsplashes, a single black-and-white textured rectangle briefly covered the change location at the moment of the change. This method allowed us to draw attention to the change location without giving away the exact nature of the change. When the change was in a central-interest element, observers were generally able to tell immediately what the change was. This confirms that the masking rectangle was not somehow wiping out the internal representation, and it shows that observers had coded the content of central-interest elements. When the change concerned a marginal-interest element, observers were often unable to determine what it was, showing

that observers had not coded the content of marginal-interest elements.

These results indicate that humans' internal representation of the visual field is much sparser than the subjective experience of 'seeing' suggests. Only the parts of the environment that observers attend to and encode as interesting are available for making comparisons. Similar inferences have been made for sensory visual memory<sup>8</sup>, but in these older experiments, simple, artificial stimuli were used, presented for only fractions of a second, whereas we presented natural visual scenes for several seconds and used very large changes.

If only attended parts of the environment are represented in the brain, how can we have the impression of such richness and completeness in the visual world outside us? The answer might be that the visual world acts as an external memory<sup>9</sup>. We have the impression of simultaneously seeing everything, because any portion of the visual field that awakens our interest is immediately available for scrutiny through an unconscious flick of the eye or of attention. However, those parts of the scene that are not being currently processed (and are in some sense not 'seen') nevertheless constitute a background or setting that enlivens our visual experience.

The idea of the world as an outside memory is receiving attention from scientists and philosophers interested in the problem of perceptual 'filling in'<sup>10</sup> of visual scotomas and illusory contours. Work in robotics with the concept of 'active' vision is also adopting the notion that using the outside world to represent information might be more efficient than making an internal copy.

**J. Kevin O'Regan\***, **Ronald A. Rensink†**,  
**James J. Clark‡**

\*Laboratoire de Psychologie Expérimentale, CNRS,  
Université René Descartes, 28 rue Serpente,  
75006 Paris, France

e-mail: oregan@ext.jussieu.fr

†Cambridge Basic Research,  
Nissan Research and Development,  
4 Cambridge Center, Cambridge,  
Massachusetts 02142, USA

‡Department of Electrical Engineering,  
McGill University, Montreal,  
Québec H3A 2A7, Canada

1. Simons, D. J. & Levin, D. T. *Trends Cogn. Sci.* **1**, 261–267 (1997).
2. Intrub, H. *Trends Cogn. Sci.* **1**, 217–221 (1997).
3. McConkie, G. W. & Currie, C. B. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 563–581 (1996).
4. Hayhoe, M. M., Bensinger, D. G. & Ballard, D. H. *Vision Res.* **38**, 125–137 (1998).
5. Rensink, R. A., O'Regan, J. K. & Clark, J. J. *Psychol. Sci.* **8**, 368–373 (1997).
6. O'Regan, J. K., Deubel, H., Clark, J. J. & Rensink, R. A. *Visual Cogn.* (in the press).
7. Levin, D. T. & Simons, D. J. *Psychonom. Bull. Rev.* **4**, 501–506 (1997).
8. Haber, R. N. *Behav. Brain Sci.* **6**, 1–54 (1983).
9. O'Regan, J. K. *Can. J. Psychol.* **46**, 461–488 (1992).
10. Pessoa, L., Thompson, E. & Noë, A. *Behav. Brain Sci.* **21**, 723–802 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

# From Daedalus to NASA

This New Ocean: The Story of the First Space Age

by William E. Burrows

Random House: 1998. 741 pp. \$34.95, £27.50

G. Siegfried Kutter

If anything distinguishes our species from others, apart from our communication skills, our ability to imagine and think abstractly and our highly developed precision grip, it must surely be our sense of adventure. The urge to explore and push the envelope of the known, either out of necessity to ensure survival or out of curiosity, is in our blood. More than 100,000 years ago, our early ancestors left their African homelands, migrated north and radiated east and west, possibly in repeated waves, to eventually settle in all continents except, until recently, Antarctica. Following these early migrations were countless others, some peaceful, but many more accompanied by violence and war. The urge to conquer and dominate took our kind back and forth across the continents and the seas. We are a restless and adventuresome species.

The journey has not ended. In fact, the ultimate journey has just begun — the journey into space. We are at a rather privileged place in the history of exploration, as Carl Sagan pointed out in a lecture in 1970: "In all the history of mankind there will be only one generation which will be the first to explore the Solar System, one generation for which, in childhood the planets are distant and indistinct disks moving through the night, and for which in old age the planets are places, diverse new worlds in the course of exploration." Very likely, our generation or that of our children will be the first to set foot on another planet in the Solar System.

This only recently acquired ability to free ourselves of the gravitational confines of Earth, set foot on the Moon and explore the Solar System is rooted in a long history. The beginnings can be traced, as far as written records reveal, to the myths and fantasies of antiquity. There is, for example, the story of Daedalus and his son, and their attempt to escape, via air, imprisonment in King Minos' labyrinth on Crete. And there are the theories of Lucian of Samosata, who, breaking with Aristotelian teachings, considered the Moon another heavenly body and wrote the first known story about a voyage to our planet's satellite in AD160. Gradually, over the course of centuries, just as astrology slowly gave way to astronomy, the ancient myths and fantasies evolved into scientific knowledge and technical know-how based on hard, demonstrable facts, laying the foundation for the development of twentieth-century rocketry and space exploration.



We have contact: Russian and US astronauts meet for the planned 1975 US–Russian spacecraft link-up.

This arduous journey from myth and fantasy to hard science and rocketry marks the beginning of *This New Ocean: The Story of the First Space Age* by William E. Burrows. In roughly the first 50 pages, Burrows traces this journey to the first modern theoreticians of rocketry, Konstantin Tsiolkovsky and Hermann Oberth, and to the more practically inclined Robert Goddard. On 16 March 1926, Goddard flew the first liquid-fuel rocket, which consisted of a small combustion chamber and nozzle connected by tubes to an oxygen tank and a gasoline tank. The rocket was 41 feet high.

In the remainder of this well-researched, clearly written, encyclopaedic treatise, some 700-plus pages in length, Burrows focuses on the twentieth century and weaves together into a cohesive story the heroic (as well as bureaucratic) efforts, the grandiose successes, the unavoidable, sometimes fatal failures and the many forces exerted by a range of interests, national and international, political and commercial, military and purely scientific, that shaped the space age.

Among many things, we learn of Germany's Second World War development of *Vergeltungswaffen* (vengeance weapons), including the V2, the true ancestor of all space launch vehicles; the competition between the United States and the Soviet Union at the close of the Second World War to capture German rocket experts and V2s to advance their respective missile programmes; the intense efforts by these two adversaries to outdo each other in all aspects of missile development during roughly four decades of Cold War competition, confrontation and paranoia; the race between

them to launch the first Earth-orbiting satellite, the first man into space and the first crew to the Moon; and the creation in 1958 of the US space agency NASA, a civilian agency given the charge, among others, to "research into problems of flight within and outside the Earth's atmosphere" and to be "devoted to peaceful purposes for the benefit of all mankind".

The book also covers the espionage and intelligence activities carried out by both adversaries. Thus, in the early 1960s, American agents 'borrowed', in the dark of night, a Russian Lunik spacecraft at a trade fair in Mexico, photographed its interior and crated it back up before it was missed. We hear of the United States' creation in 1960 of the National Reconnaissance Office (originally named the Office of Missile and Satellite Systems), which was charged to buy and operate the US spy satellites, but whose very existence was officially denied until 1992. And there is the contentious history of the US Space Transportation System and space shuttle programme, which early on pitted NASA and its congressional supporters against the Air Force, the Office of Management and Budget, the General Accounting Office and industry, which had a stake in expendable launch vehicles.

Burrows discusses many of the space science missions, from Lunik and Pioneer 4, launched in 1959 (two lunar probes that missed their target and instead became the Sun's first artificial 'planets'), Mariner 2 (the first spacecraft to reach and speed past another planet, Venus, in 1962), and Mars 2 and 3 (the first spacecraft to attempt soft landings on the Red Planet, in 1971, only to



be foiled by the worst martian dust storm in recorded history), to the many dozens of missions launched since, or currently being developed, by NASA, the European Space Agency, Japan, Russia and other nations.

With the disintegration of the Soviet Union in 1991 and the end of the Cold War, the 'First Space Age', as defined by Burrows, ended and the 'Second Space Age' began. The competition between the United States and the Soviet Union was replaced by growing international cooperation, as well as by increasingly intense commercial competition, in which the former rivals were joined by China, France, India, Israel and Japan. Commercial communications satellites proliferated. Global positioning, sophisticated remote-sensing capabilities and intense competition in the development of cheap access to space spread into the private sector. NASA, in partnership with agencies from 15 nations, including Russia, began a concerted effort to build the International Space Station. And all nations found it necessary to place strict limits on spending on new space science missions.

At the same time, following the Persian Gulf War in 1990–91 (which the US Air Force called the 'first space war'), the Pentagon began to consider space as the 'fourth medium', after land, sea and air. This point was made clear to the US Senate Armed Services Committee in 1991 by General Donald Kutyna, head of the US Space Command: "Unless we have a sound space-control capability, we may find ourselves in a conflict with a nation with space forces while we have no means to prevent space-supported attacks on ourselves and our allies." To meet this challenge, and to make the nation's aerospace industry more competitive with other space-faring nations, the United States began developing anti-satellite weapons and expanding its reconnaissance and surveillance capabilities. Burrows devotes the final chapter of *This New Ocean* to these developments.

The frontiers of human exploration have changed from those of 100,000 years ago, but the spirit and motivations have not. We are still pushing the envelope, perhaps more than ever, and certainly with more know-how and finesse. Survival — economic, national and ideological — is still the main engine. Curiosity and excitement are alive and well also, as we probe the farthest reaches of the cosmos, search for the origin of galaxies and galaxy clusters, stars and planets (solar and extrasolar), study Earth from space, assemble the International Space Station and debate setting up an outpost on Mars. Our species may have its failings, but we are not dull. Burrows makes this point convincingly.

In writing *This New Ocean*, Burrows drew on years of experience as a former aviation and space reporter for *The New York Times*, *The Washington Post*, *The Wall Street Journal*

and other publications. The book will make absorbing reading for anyone truly interested in space exploration. It could be used as a general reference on the subject (the index is thorough and well laid out), and it would serve well as a text for advanced university courses on the history of space exploration. I recommend it highly. □

G. Siegfried Kutter is at TRW, Inc., PO Box 6759, Breckenridge, Colorado 80424, USA.

## Counting on statistics

**The Politics of Large Numbers: A History of Statistical Reasoning** by Alain Desrosières (translated by Camille Naish)  
*Harvard University Press: 1998. 368 pp. \$45, £27.95*

Jonathan Rosenhead

As I write, the content of today's newspaper (like any other day's paper) depends on numbers. There are articles on radio listening figures, an epidemiological study of child allergy risks, a European Union decision on phasing out imperial measures, evidence on race bias in senior civil service posts, statistics on film audiences. And that is excluding the financial pages. We live in a world in which our understanding and our debates are supported by dense and interlocking networks of indices and percentages.

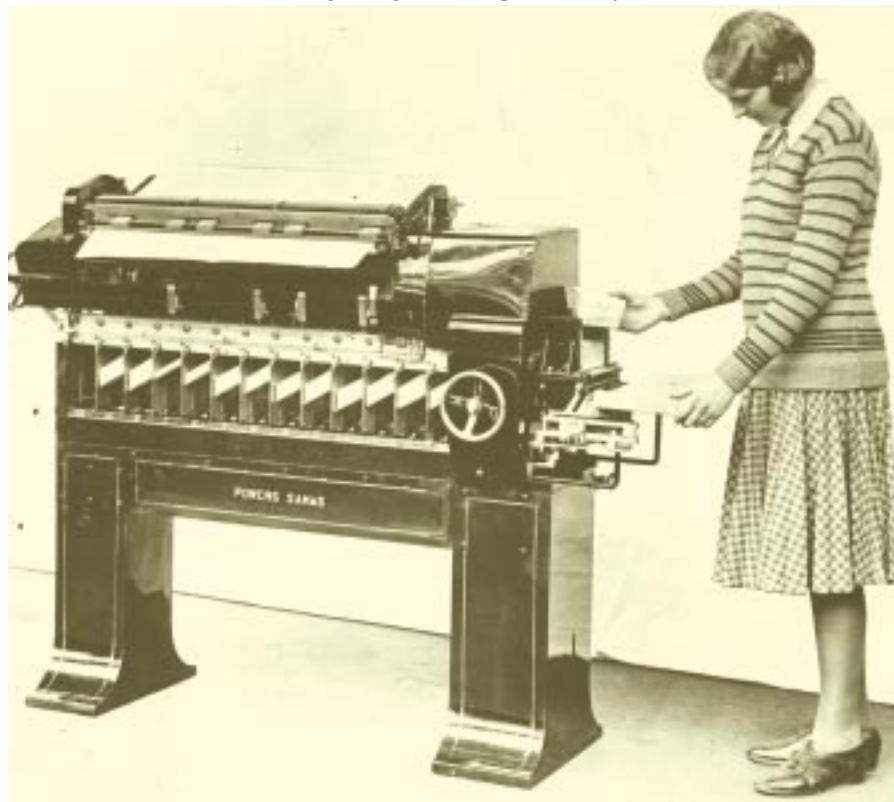
It was not ever thus. In the beginning the

public sphere, itself an evolving category, was an almost number-free zone. It has taken centuries of often ill-tempered discussion to generate the conventions that now govern what to measure, how to aggregate, how to interpret deviations, what policies and decisions require numerical backing. It is this intellectual journey that is the topic of Alain Desrosières' impressive book.

Statistics, with its aura of dispassionate dustiness, does not have a good image. It is detested by generations of social-science students, a grim necessity for medical researchers, distrusted by the general public. Many of these — and some statisticians — would be surprised to discover how often statistics has responded to social developments or even influenced them.

The broad theme of the book is that statistical measures and probabilistic concepts are most usefully seen as matters of convention, rather than of objective reality. The social context generates the need to make things countable and to interpret the counts; it also conditions the conventions that emerge. Statistics has thus engaged in a process of creating factual 'objects' accessible to common knowledge. These facts then acquire a reality because of the social practices that congeal around them.

This social process of 'objectification' has operated both at the level of statistical tools and at the level of applications. An example of the tools is the development of confidence intervals early this century: this "made imprecision and margin of error into a respectable object" rather than a matter for



Lies, damn lies and statistics: the machine responsible for processing the results of the 1931 census.

CORBIS/HULTON-DEUTSCH

shameful silence. As an example of application, consider the case of unemployment. To determine the unemployment rate, it was necessary to make a whole range of judgments. Where did people who worked only intermittently, when they got the chance, fit? Or those who through discouragement were no longer trying to find a job? Or people in poor physical or mental health? The conventions to be used in locating the hazy boundary between the potentially economically active and inactive populations were hotly argued, provoked especially by the US census of 1930. Depending on the conventions adopted in these doubtful cases, millions of unemployed could be summoned up or dispersed. However, once established, the measure could provide structure to debates on public policy that were otherwise close to being undecidable. In turn, it was those debates, in the depth of the depressed 1930s, that motivated the search for an acceptable measure. And it was the significance this measure achieved that tempted later governments to operate on social reality by tinkering with the way it was computed.

Behind this broad theme lies Desrosières' scholarly dissection of the entangled heritage of what has only recently become (almost) one subject. We are led back to the nominalism of William of Occam, and sideways to the contrast between the Linnean system and Buffon's method for classifying species. We are taken into bypassed debates, and introduced to forgotten precursors. It is not an easy read. However, the density is relieved both by fascinating nuggets of unexpected information and by accounts of set-piece confrontations.

Thus, statisticians can learn where those urns that are always cropping up in texts on probability were first manufactured. They can find out about the astronomical roots of the least squares method and why contingency tables are so called. A more general readership can pick up the motivation for Charles Booth's poverty survey (to determine the scale of the 'very poor' population, with a view to expelling them from London) and learn how slaves were treated in the head-count on which US congressional apportionment was based (answer: as three-fifths of a free man). Much of the technical repertoire of mathematical statistics was developed by British eugenicists as weapons for their 'scientifico-political war machine' directed against all policies that might encourage the poor to breed.

I have three reservations about this book. One is its Gallocentricity. Most of the bibliography is in French, and a novice in the field might form a distorted impression of the importance of French administrators and probabilists. The less than satisfactory translation too often displays infelicities, along with a lack of awareness of English idiom and even some technical terminology.

The second reservation is that the delay in publication (the French publication date was 1993) has made a difference. Surprisingly, given the book's broad historical scope, a good deal has happened. There is the escalation of interest and theorizing on risk and the 'risk society', and the publication of Peter L. Bernstein's major work, *Against the Gods* (John Wiley, 1998). And there is the all-out assault by Alan Sokal and Jean Bricmont (in their book *Intellectual Impostures*, Profile, 1998) on a range of French philosophers and sociologists who deal with science. One of their identified targets is the work of Bruno Latour, with its emphasis on 'science in the making'. Desrosières cites Latour as an influence, and his book is indeed an excellent example of the insights that the approach can offer. So the absence of a riposte to the Sokal attack, though inevitable, is nevertheless frustrating.

Finally, there is the title: "Politics"? "The Political Sociology of Large Numbers" is about as far as I would go. If you want the politics of large numbers, then go instead for *Statistics in Society: The Arithmetic of Politics*, published by Arnold in 1998, and edited by Daniel Dorling and Stephen Simpson on behalf of the Radical Statistics Group. □

Jonathan Rosenhead is in the Department of Operational Research, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK.

## Decision-making for our ecological future

### Conservation of Biological Resources

by E. J. Milner-Gulland and Ruth Mace  
*Blackwell Scientific: 1998. 424 pp.*  
£24.95, \$59.95

Marc Mangel

The human footprint will soon be felt almost everywhere on the globe. And although the term 'ecosystem management' is bandied about these days, the reality is that we do not manage ecosystems; rather, we manage human intervention in ecosystems. Thus, it is incumbent on those interested in biodiversity conservation to understand the economics and behaviour associated with the interaction between people and the biotic world. *Conservation of Biological Resources* will allow us to do this.

E. J. Milner-Gulland and Ruth Mace start by laying out the problem of conservation through a brief historical review and a summary of initiatives taken on sustainable development by the IUCN (the World Conservation Union). There then follow two substantive sections on the theoretical background and on case studies, with a concluding section on setting conservation to work.

The authors emphasize the importance of understanding the ecological and social issues affecting biological resources. To be sure, some people will be put off by the notion of 'biological resources', arguing that it is this notion itself that causes lack of conservation. However, as the authors note, "most wild species are found in areas where people either live or venture to harvest resources", and resource use and its effects are surely inevitable. To shy away from important concepts such as harvesting theory because they are somehow tainted by use is to bury one's head in the sand. Important conservation problems will not be solved by ignoring them.

The book's coverage of the ecological and economic theory of sustainable use, decision-making and practical considerations when applying the theory is beautifully done and very readable. Although no prior knowledge of the field is needed, readers with some calculus and population ecology will find it easier going.

The authors stress the importance of the theoretical material, and I concur: ecological and combined ecological-economic theory illuminates many of the issues associated with conservation in ways that less precise approaches cannot do. Similarly, they emphasize the importance of considering trade offs and uncertainty in decision-making. Too often, conservation biologists have refused to discuss trade offs (in which case someone else makes them), or to act in the face of uncertainty, arguing that more data are needed (in which case someone else acts). The theory of decision-making provides a framework for dealing with both trade offs and uncertainty.

The choice of case studies in the book is excellent, with topics including sustainable forestry, wildlife management, squid fisheries and eco-tourism. I would have liked an introduction to each case study, explaining why it was picked and highlighting ideas from the book's second part that are involved in the case study. Finally, I am concerned that the concluding discussion on dealing with institutions, economics and politics to make conservation work will get lost in the larger text. Nevertheless, this fine volume is sure to be important and will set a standard for truly interdisciplinary work in conservation biology. □

Marc Mangel is in the Department of Environmental Studies, University of California, Santa Cruz, California 95064, USA.

More along these lines

### **Biodiversity Dynamics: Turnover of Populations, Taxa and Communities**

edited by Michael L. McKinney  
& James A. Drake. *Columbia University Press:*  
1998. \$69, £48

### **Ecosystem Management: Adaptive Strategies for Natural Resources Organizations in the 21st Century**

edited by Jennifer Aley, William R. Burch,  
Beth Conover & Donald Field. *Taylor & Francis:*  
1999. £20.95, \$24.95

## Behavioral Ecology and Conservation Biology

edited by Tim Caro. *Oxford University Press:* 1998.  
£65, \$90

## Super news

Superconductivity and Superfluidity  
by Toshihiko Tsuneto (translated by  
Mikio Nakahara)  
*Cambridge University Press:* 1998. 207 pp.  
£45, \$69.95

A. J. Leggett

Most of today's standard textbooks on superconductivity (and on the superfluidity of liquid  $^4\text{He}$ ) date from the heyday of research on 'low-temperature' superconductors in the 1960s. But since then, in addition to these 'classic' superfluids, several related but interestingly different superfluid systems have been realized experimentally — the low-temperature phases of liquid  $^3\text{He}$ , the 'heavy-fermion' and copper oxide superconductors and the (presumably superfluid!) Bose-condensed phase of dilute alkali gases.

*Superconductivity and Superfluidity*, which originally appeared in Japanese in 1993, is intended in some sense as an update of, or supplement to, the standard texts in the light of these developments. However, it concentrates mainly on the Fermi superfluids, with a couple of short sections on  $^4\text{He}$  and a postscript on the alkalis.

A distinguishing feature of the book is its treatment of Cooper pairing, which starts from an arbitrary spin and orbital state, thereby permitting unified treatments of classic superconductivity, copper oxide (and heavy-fermion) superconductors and  $^3\text{He}$ . Although the coverage of these systems is not complete (impossible in a book of this length), most of the key phenomena, at both the microscopic and the phenomenological level, are addressed.

The exposition is well organized and crisp, and the translation generally runs smoothly. The book should be useful to graduate students and those who have had some exposure to field-theoretic and diagrammatic methods but who are not ready to take the plunge into more specialized texts such as that of Vollhardt and Wölfle on superfluid  $^3\text{He}$ . □

A. J. Leggett is at the Loomis Laboratory of Physics,  
University of Illinois, Urbana, Illinois 61801-3080,  
USA.

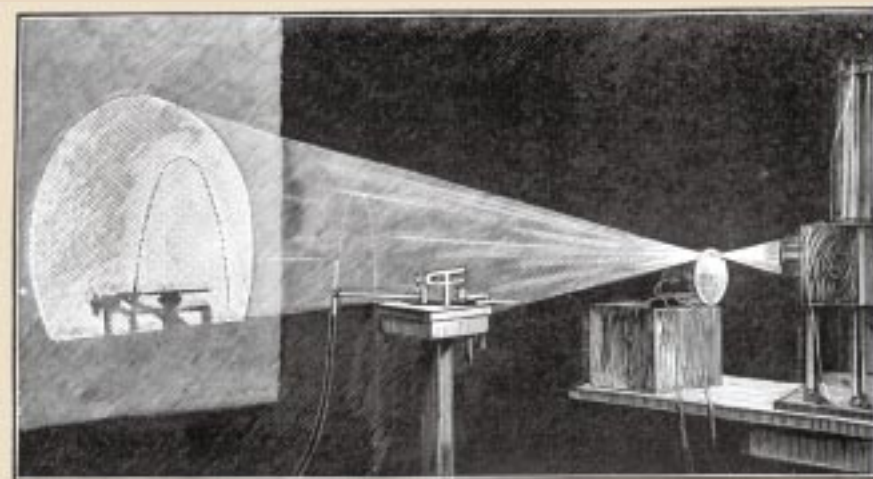
Also available

## Superconductivity

J. B. Ketterson & S. N. Song

*Cambridge University Press:* 1999. £29.95, \$49.95

## Science in culture



### Boys's bubbles

Martin Kemp

The burgeoning shelves of 'popular science' in bookshops seem to suggest that we are witnessing a recent phenomenon. In fact, campaigns to broadcast science to mass audiences proliferated in the nineteenth century, and had been preceded by such individual initiatives as Count Francesco Algarotti's *Newtonianism for Ladies* in 1737. In Britain, luminaries such as the polemical physicist Sir David Brewster aspired through accessible publications and the founding of the British Association in 1831 to educate the public in the observational wonders of natural phenomena and experimentation, above all in a Christian spirit of devotion to nature.

One of the most delightful and enduring contributions was Sir Charles Vernon Boys's *Soap Bubbles and the Forces Which Mould Them*, originally delivered as three lectures for young people at the London Institution during the winter of 1889–90, and first published by the Society for Promoting Christian Knowledge in 1890. As late as 1959 they were reprinted in the American "Science Study Series". John Durston remarked in his preface, "that this book remains as valuable to beginning scientists today as the lectures were seventy years ago is remarkable in this age of revolution".

Boys, born in 1855 as the son of a clergyman, studied chemistry and physics at the Royal School of Mines. His first paper, published in 1880 under the wing of T. H. Huxley, described experiments to determine whether spiders would be lured by the 'buzz' of a tuning fork — which they were. Renowned for having determined the Newtonian constant for gravitation, he was also the inventor of quartz fibres, which he drew out by shooting a crossbow, and an innovator in instantaneous photography. He was delighted to discover that a quartz fibre dipped into castor oil was indistinguishable in micro-photographs from the beaded filaments made by a spider. Music, fibres, viscous beads, photography and spiders became interwoven themes in his rapturous exploration of

C. V. Boys's "Experiment for showing by intermittent light the apparently stationary drops into which a fountain is broken up by the action of musical sound", from *Soap Bubbles and the Forces Which Mould Them*, 1890.

the visual and aural music of physical phenomena.

Typical of his approach is the experiment that provided the frontispiece for his little volume on soap bubbles. A beam of light is projected onto a screen through a small hole in a card, behind which is a spinning disk with six holes round its rim. The pulsing beam casts the shadow of a fountain which is vibrated by a tuning-fork. The speed of the disk is literally fine-tuned by blowing through the holes until the note is precisely that emitted by the tuning-fork.

The apparently continuous jet is revealed as an arc of beads. If the card turns fractionally more slowly, "all the drops will appear to slowly march onwards, and what is so beautiful ... each little drop may be seen to gradually break off, pulling out a waist which becomes a little drop, and then when the main drop is free it slowly oscillates, becoming wide and long, or turning over and over, as it goes on its way".

This pretty experiment is followed by an account of Chichester Bell's 'Singing water jet', in which an arched fountain falling on a rubber sheet stretched over the end of a tube "begins to sing of its own accord" under the agitation of vibrations transmitted to its nozzle from a chiming watch and muffled musical box.

In his love of the visual realization of musical sounds through the exciting of vibrations in liquids, Boys stood in a distinguished tradition that includes Christopher Wren and Robert Hooke. They would not in any way have dissented from Boys's conviction that "experiment is a question we ask of Nature, who is always ready to give a correct answer, provided we ask properly, that is, provided we arrange a proper experiment". □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.



# Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport

Ingrid R. Vetter, Christine Nowak, Takeharu Nishimoto\*, Jürgen Kuhlmann & Alfred Wittinghofer

Abteilung Strukturelle Biologie, Max-Planck-Institut für molekulare Physiologie, Postfach 102664, 44026 Dortmund, Germany

\* Department of Molecular Biology, Graduate School of Medical Science, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-82, Japan

**The protein Ran is a small GTP-binding protein that binds to two types of effector inside the cell: Ran-binding proteins, which have a role in terminating export processes from the nucleus to the cytoplasm, and importin- $\beta$ -like molecules that bind cargo proteins during nuclear transport. The Ran-binding domain is a conserved sequence motif found in several proteins that participate in these transport processes. The Ran-binding protein RanBP2 contains four of these domains and constitutes a large part of the cytoplasmic fibrils that extend from the nuclear-pore complex. The structure of Ran bound to a non-hydrolysable GTP analogue (Ran·GppNHp) in complex with the first Ran-binding domain (RanBD1) of human RanBP2 reveals not only that RanBD1 has a pleckstrin-homology domain fold, but also that the switch-I region of Ran·GppNHp resembles the canonical Ras·GppNHp structure and that the carboxy terminus of Ran is wrapped around RanBD1, contacting a basic patch on RanBD1 through its acidic end. This molecular ‘embrace’ enables RanBDs to sequester the Ran carboxy terminus, triggering the dissociation of Ran·GTP from importin- $\beta$ -related transport factors and facilitating GTP hydrolysis by the GTPase-activating protein ranGAP. Such a mechanism represents a new type of switch mechanism and regulatory protein–protein interaction for a Ras-related protein.**

The superfamily of small Ras-like GTP-binding proteins is involved in many regulatory processes in the cell: for example, Ras regulates cell proliferation by switching between an active GTP-bound form and an inactive GDP-bound form; a similar mechanism is used by Arf to regulate vesicle transport processes, by Rho, Rac and Cdc42 in cytoskeleton rearrangements and by Ran to achieve directionality in nuclear–cytoplasmic transport processes (for a review on Ran, see ref. 1). Effectors of Ras-related GTP-binding proteins are proteins that bind tightly to the GTP-bound form but only weakly to the GDP-bound form. Ran-binding effectors are Ran-binding proteins (RanBPs)<sup>2,3</sup>, which are putative nuclear-export terminators, and importin- $\beta$ -like (also called karyopherin- $\beta$ -like) molecules, which bind import and export cargo proteins. The group of importin- $\beta$ -like molecules comprises both import- and export-related transport proteins, the latter being called exportins. Complexes of importin- $\beta$  and cargo protein are dissociated by Ran·GTP when they enter the nucleus<sup>4,5</sup>, whereas export complexes of exportin and cargo proteins can only be formed in the nucleus with the participation of Ran·GTP (ref. 6). Binding of importin-like molecules to Ran inhibits GTP hydrolysis stimulated by ranGAP (where GAP is GTPase-activating protein)<sup>5,7</sup>.

In contrast to importin- $\beta$  like proteins, RanBP-like effectors act as co-activators of ranGAP and essential release factors for complexes of Ran with transport factors<sup>8,9</sup>. RanBP2 is a large protein located on the cytoplasmic side of the nuclear-pore complex (NPC) and is essential for protein import and/or export in higher eukaryotic cells<sup>10,11</sup>. The 3,224 residues of RanBP2 comprise an amino-terminal 700-residue leucine-rich region, four Ran-binding domains (homologous to RanBP1), eight zinc-finger motifs, and a carboxy terminus with high homology to cyclophilin, a protein of unknown function with peptidyl-prolyl isomerase activity<sup>12,13</sup>. RanBP2 contains the XFXFG pentapeptide repeats characteristic of NPC proteins that may be involved in docking transport complexes to nucleoporins<sup>14</sup>. RanBP2 is a filamentous molecule of

~36 nm in length that forms a major part of the cytoplasmic fibrils of the NPC<sup>11</sup>; in rat-liver nuclear envelopes, it is the only prominent protein that binds Ran·GTP (ref. 15). It has been proposed that Ran targets members of the importin- $\beta$ /karyopherin- $\beta$  superfamily to the NPC by binding to the Ran-binding domains of RanBP2 (ref. 16). Here we report the crystal structure of the complex of the first Ran-binding domain (RanBD1) of human RanBP2 with Ran bound to a non-hydrolysable GTP analogue (GppNHp). The structure suggests a molecular mechanism to explain the role of RanBPs in nuclear transport.

## Crystallization and structure determination

We used a RanBD1 construct containing 167 residues corresponding to residues 1,155–1,321 of the original human RanBP2 sequence<sup>12,13</sup>, which we number as residues 1 to 167 for simplicity. To distinguish RanBD1 residues from Ran residues, the latter are identified by a superscript, so 181<sup>ran</sup> means residue 181 in the Ran moiety. Monoclinic crystals of the complex of RanBD1 with Ran·GppNHp diffract to 2.9 Å at the synchrotron. The structure of the complex was determined by molecular replacement using Ran·GDP (ref. 17) as a model. The electron density was averaged between two molecules in the asymmetric unit and allowed unambiguous tracing and building of the missing RanBD1 moiety, as well as of the rearranged C terminus and switch-I region (effector loop) of Ran·GppNHp. A representative section of the map encompassing residues 25–36<sup>ran</sup> of the switch-I region of Ran·GppNHp and part of the conserved <sup>57</sup>WKER-region of RanBD1 (Glu 59 and Arg 60) is shown in Fig. 1. The seven N-terminal residues (1–7<sup>ran</sup>) of both Ran molecules are disordered, as are the last five residues (212–216<sup>ran</sup>) of the first Ran molecule. As the density for residues 188 to 216 of the second Ran molecule was very weak, they were omitted from the model. No electron density is visible for the N-terminal end of the RanBD1 construct, so residues 1–16 cannot be traced in either molecule. The C terminus of both RanBD1 domains is

disordered from residue 151. Thus, the model comprises residues 8–211 of the first Ran-GppNHp molecule, residues 8–187 of the second, and residues 17–150 of the two RanBD1 domains. Crystallographic data are summarized in Table 1.

### The Ran-triphosphate conformation

The structure of Ran bound to GDP revealed a large conformational deviation in the switch-I region compared with the GDP-bound forms of Ras, Rap, Rho and Cdc42, but similar to the GDP-bound form of Arf (refs 18, 19) and EF-Tu (refs 20, 21). Unlike the effector loop that has no regular secondary structure in the canonical G-domain structure of Ras, Ran-GDP has an extra  $\beta$ -strand in the effector region, termed  $\beta_{2E}$ . This different conformation of the effector region in Ran-GDP puts two residues that are highly conserved in Ras proteins, a phenylalanine (Phe 35<sup>Ran</sup>) and an invariant threonine (Thr 42<sup>Ran</sup>), far away from the nucleotide-binding pocket (Fig. 2a). Another difference between Ran and other members of the Ras superfamily is the presence of a long C-terminal extension in Ran, which consists of a linker without secondary structure and then a 16-residue  $\alpha$ -helix, situated opposite the switch-I region and ending in a highly acidic C-terminal DEDDDL motif (residues 211–216<sup>Ran</sup>) whose charge is conserved in all Ran proteins and which showed no electron density in the original Ran-GDP structure<sup>17</sup>.

In the crystal structure we have determined for the complex, Ran-GppNHp adopts a different conformation: the most pronounced changes are in the switch-I region (Fig. 2a) and in the C-terminal end (residues 173–216<sup>Ran</sup>, Fig. 2b). The two extra helical turns of  $\alpha_{1b}$  and the extra  $\beta$ -strand in switch I have melted so that the protein assumes a conformation similar to the canonical triphosphate structure of Ras-GppNHp (Fig. 2b) and of all other related GTP-binding proteins whose structures are known. The main-chain NH of Thr 42<sup>Ran</sup>, corresponding to Thr 35 in Ras, hydrogen-bonds to the  $\gamma$ -phosphate and its side-chain hydroxyl is coordinated to Mg<sup>2+</sup>. This large 14-Å movement of Thr 42<sup>Ran</sup> and its coordination were predicted from the Ran-GDP structure<sup>17</sup> and, together with the coordination of Gly 68<sup>Ran</sup> (Gly 60 in Ras), is considered to be the basis of the universal conformational change in GTP-binding proteins<sup>22</sup>. There is a similar conformational

**Table 1 Crystallographic analysis**

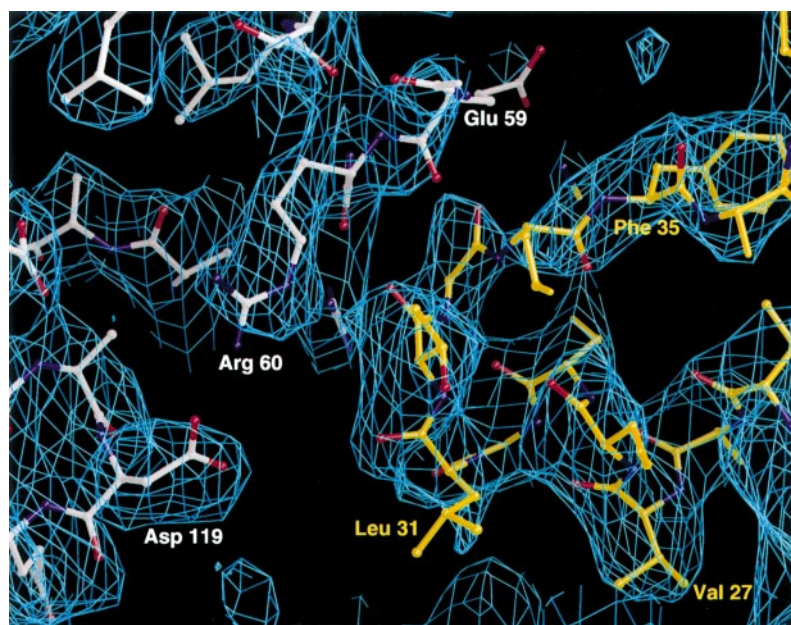
|                                         |                                   |
|-----------------------------------------|-----------------------------------|
| Resolution (Å)                          | 30–2.9                            |
| Observed reflections                    | 90,705                            |
| Unique reflections                      | 20,286                            |
| Completeness                            | 86% to 2.9 Å (66% from 3.0–2.9 Å) |
| $I/\sigma$                              | 13.1 (3.1 from 3.0–2.9)           |
| $R_{\text{merge}}^*$                    | 9.1%                              |
| Refinement                              |                                   |
| Resolution (Å)                          | 12.5–2.96                         |
| Reflections                             | 17,709                            |
| $R_{\text{cryst}}^\dagger$              | 25.2%                             |
| $R_{\text{free}}$                       | 30.4%                             |
| Model                                   |                                   |
| Non-hydrogen atoms:                     |                                   |
| Protein                                 | 5,302                             |
| Magnesium ions                          | 2                                 |
| Nucleotide                              | 65                                |
| Water                                   | 36                                |
| R.m.s. deviation from expected geometry |                                   |
| Bond lengths (Å)                        | 0.008                             |
| Bond angles (deg)                       | 1.41                              |
| Overall $B$ -value (Å <sup>2</sup> )    | 56.1                              |

\*  $R_{\text{merge}} = \sum_i \sum_l |I_i - \langle I \rangle| / \sum_i \sum_l I_i$ , where  $I_i$  is the intensity for the  $i$ th measurement of an equivalent reflection with indices  $h, k, l$ .

†  $R_{\text{cryst}} = \sum_i |F_{\text{obs}} - F_{\text{calc}}| / \sum_i F_{\text{obs}}$ , where  $F_{\text{obs}}$  denotes the observed structure factor amplitude, and  $F_{\text{calc}}$  denotes the structure factor amplitude calculated from the model; 10% of reflections were used to calculate  $R_{\text{free}}$ .

change in the transition of EF-Tu-GDP to EF-Tu-GppNHp (refs 20, 21) and in Arf (ref. 23) which, like Ran, have an extra  $\beta$ -sheet in switch I in the GDP-bound form<sup>18,19</sup>.

The second important conformational change involves the C-terminal extension of Ran (that is, a C-terminal switch), which adopts a completely different conformation but retains the secondary structure of the helical segment (Fig. 2b). This is surprising because the C terminus does not contact the nucleotide in either the GDP- or the GppNHp-bound form. Overlaying the structures of Ran-GDP and Ran-GppNHp (Fig. 2c, d) reveals that the triphosphate conformation of the effector loop is not feasible in the GDP form of Ran because residues 33–35<sup>Ran</sup> of the effector region would clash with residues 182<sup>Ran</sup>, 183<sup>Ran</sup> and 186<sup>Ran</sup> in the C-terminal linker, suggesting that this stretch is displaced when GTP is bound owing to steric hindrance. This



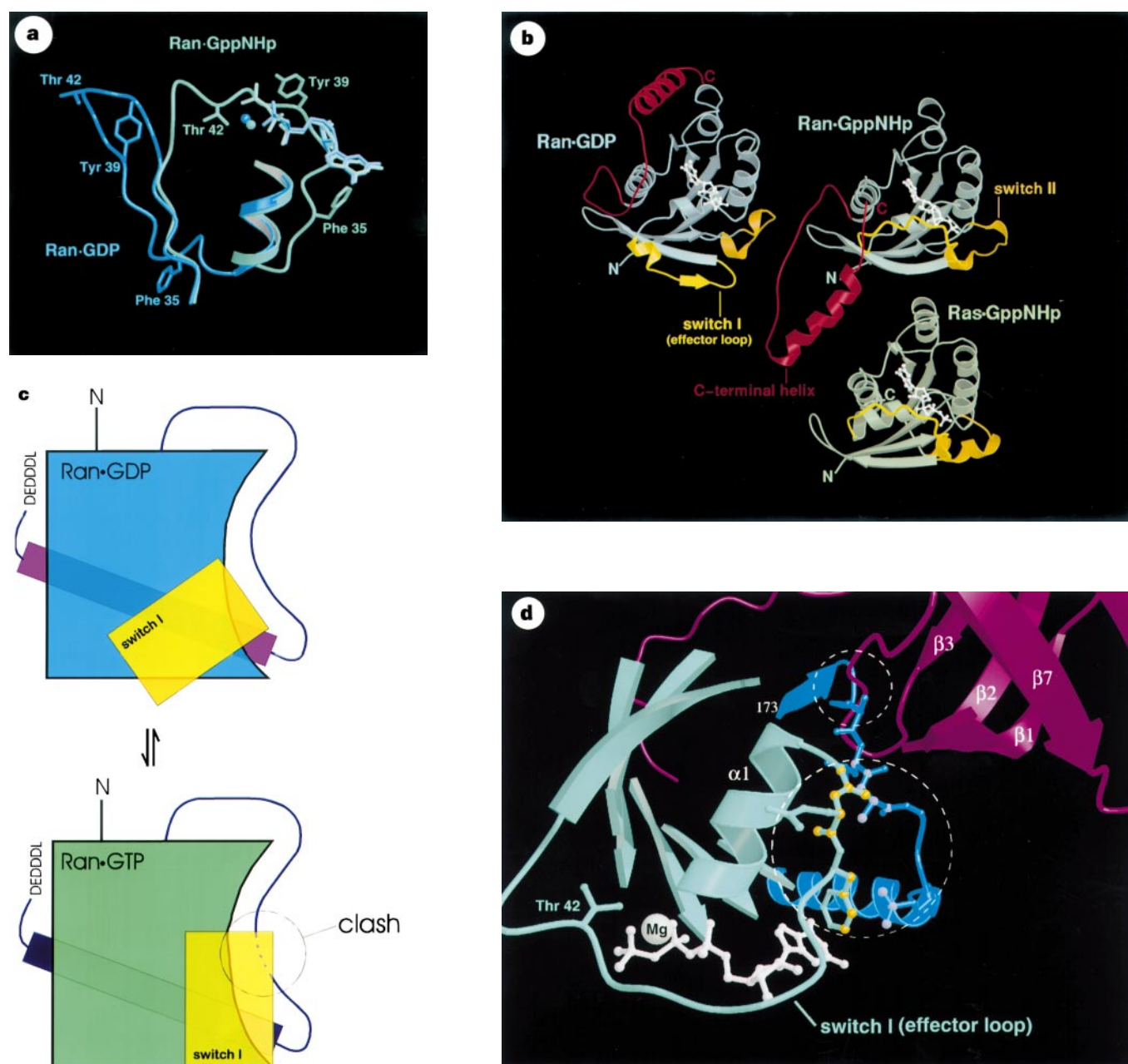
**Figure 1** Representative electron density around switch I in Ran-GppNHp and the conserved WKER motif of RanBD1 (residues 57–60). Residues from RanBD1 are

represented by white carbon traces, residues from Ran by yellow carbon traces. The omit map is contoured at 1.1  $\sigma$ . The figure was prepared with BOBSCRIPT<sup>46</sup>.

destabilization of the linker may cause dislocation of the C-terminal helix of Ran-GTP or at least loosening of its interaction with the rest of the protein. The C terminus might be involved in a similar structural change in Ran-GTP alone, which would be in line with experimental observations. Antibodies raised against a peptide sequence corresponding to the residues immediately preceding the C-terminal acidic DEDDDL motif of Ran can immunoprecipitate Ran in the GTP form, but not in the GDP form<sup>24</sup>, indicating that this sequence could be fixed in the latter but accessible in the former.

The DEDDDL motif presumably fixes the Ran C-terminal helix

by interacting with a basic patch (residues 139–142<sup>Ran</sup>) on Ran-GDP (ref. 17). This motif stabilizes the GDP-bound form and destabilizes the GTP-bound form<sup>24</sup>. The crystal structure suggests that deletion of the DEDDDL motif could stabilize the GTP-bound state by loosening the interaction of the C-terminal end with the core G-domain of Ran, allowing the linker region to adopt a conformation that does not interfere with the effector loop in its GTP-bound conformation (Fig. 2b–d). The acidic sequence presumably stabilizes the GDP-bound form by interacting with the basic patch on the surface of Ran-GDP, keeping the linker region in



**Figure 2** The Ran-GTP-analogue structure and conformational change. **a**, Conformational change of the effector loop (switch I region) between the Ran-GDP (cyan) and Ran-GppNHp (green) form. Residues relevant to the switch mechanism are coloured according to their conformational state. Spheres depict the position of magnesium ion in the two conformations: it is contacted by Thr 42 in the triphosphate conformation. **b**, Ribbon representation of the structures of free Ran-GDP (ref. 17), Ran-GppNHp from the RanBD1 complex, and Ras-GppNHp (ref. 47). The important elements of the conformational switch are shown.

**c**, Mechanism of the C-terminal switch mechanism; **d**, detailed view of the steric clash (larger dashed circle) between the effector region and the C-terminal linker. Ran-GppNHp structural elements are shown in green and the linker region of Ran-GDP in cyan; GppNHp is in white and RanBD1 in magenta. The side-chain atoms that clash are picked out in yellow (effector region) and lilac (C-terminal linker). The figure was generated using MOLSCRIPT and Raster3D (refs 48, 49). The smaller dashed circle indicates a potential steric clash between Ran-GDP and RanBD1.



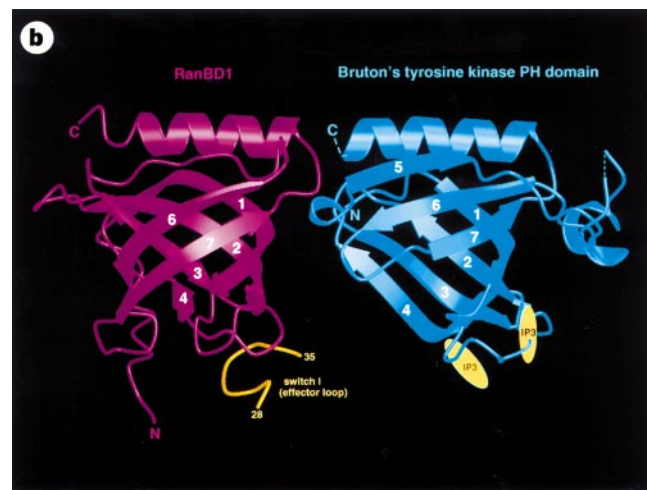
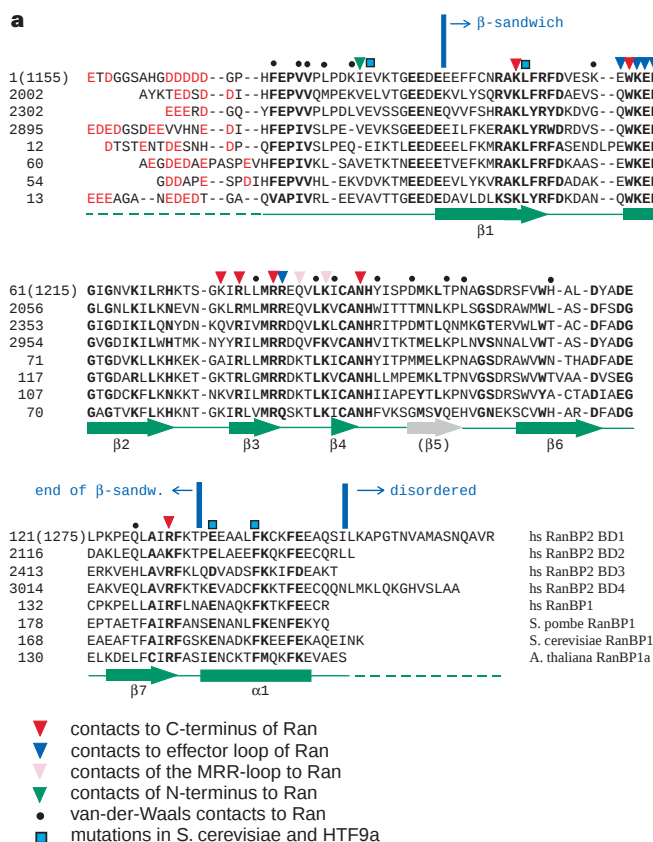
place and hindering the effector loop from assuming the GTP-bound conformation. This might account for the tenfold decrease in affinity of Ran for GTP compared with GDP, whereas other Ras-like proteins have a higher affinity for GTP (ref. 25).

The change in effector-loop conformation also explains why NTF2, a nuclear-transport factor that mediates the nuclear import of Ran (ref. 26), binds to the GDP but not the GTP form of Ran. Superposition of the Ran moieties in the Ran-GDP-NTF2 (ref. 27) and the RanBD1-Ran-GppNHp complexes shows that the effector loop of Ran-GppNHp would clash with residues 93–95 of NTF2. Also, the switch-II region of Ran has a different conformation in the two complexes and is intimately involved in the contact between Ran-GDP and NTF2, but as this part of the molecule is flexible, the effector loop is probably the crucial element that prevents binding of NTF2 to Ran-GppNHp.

### RanBD1 as a PH/PTB-domain protein

The residues in the RanBD1 structure that can be identified in the electron density constitute the minimal domain that can be expressed in *Escherichia coli* and can still bind Ran-GTP (Fig. 3a; ref. 28). Secondary-structure predictions using several different algorithms (for example, GOR, PHD, SIMPA96, SOPM, PREDATOR) all show that RanBD1 consists almost entirely of  $\alpha$ -helices. Unexpectedly, the RanBD1 domain has the  $\beta$ -barrel topology of a pleckstrin-homology (PH) and the similar phosphotyrosine-binding (PTB) domain, although the sequence shows no detectable homology to these protein domains<sup>29</sup>. This  $\beta$ -barrel fold has also been called an orthogonal sandwich, or an up-and-down  $\beta$ -barrel<sup>29</sup>. Thus, the RanBD1 structure features two almost orthogonal  $\beta$ -sheets consisting of four and two strands, with a single helix sitting on top of the sheet like a lid (Fig. 3b). There is a slight deviation for the PH-domain fold in that RanBD1 contains only two strands in the second sheet instead of three (Fig. 3b). Structural overlays show that RanBD1 is more closely related to PH domains, the closest homologue being the PH domain of Bruton's tyrosine kinase<sup>30</sup>, with an r.m.s. deviation of 1.4 Å for 100 C $\alpha$  atoms (Fig. 3b). PH domains are found in several proteins involved in the regulation of signal transduction by small GTP-binding proteins, such as the Ras-specific p120GAP, the rasGEF Sos, and all of the rhoGEFs (GEF, guanine-nucleotide-exchange factor). PH domains bind phosphatidylinositol lipids and the structures of two such complexes have been solved (Fig. 3b). The PH domain may be a mediator of protein-protein interaction, given that the  $\beta$ -adrenergic-receptor kinase ( $\beta$ ARK) can bind the  $\beta\gamma$  subunits of G proteins, although the participating residues responsible only partly overlap with the PH domain<sup>31</sup>. We have demonstrated that a PH-like domain can bind to a small GTP-binding protein, showing that the PH-domain fold is a stable and versatile domain.

Sequence comparison (Fig. 3a) reveals several highly conserved regions in RanBDs, two of which, <sup>61</sup>GIGNVKIL and <sup>140</sup>FKCKFEE, are not involved in Ran binding; this sequence conservation can be explained by structural constraints. Two temperature-sensitive mutants of the yeast homologue of RanBD1, *yrb1*, have been characterized (Fig. 3a; ref. 32). The phenotype exhibits defects in nuclear transport and in the interaction with the yeast homologue of Ran, Gsp1. The mutations in the *Saccharomyces cerevisiae* homologue are L93P from the highly invariant <sup>44</sup>RAKLFRED motif (mutated residue, L), and a double mutation, F187S from the <sup>140</sup>FKxxFEx motif (first F mutated) and E182D located very close to it (Fig. 3a). The L93P and the F187S mutations are both in the hydrophobic core of RanBD1 and so would destabilize the protein. Residue E135 (E182 in *S. cerevisiae*) points towards the outside at the start of the helix, forming a 'lid' on top of the  $\beta$ -barrel, and is unlikely to cause temperature sensitivity in the double mutant. A highly conserved region is the <sup>89</sup>CANHYI sequence that ends  $\beta$ -sheet 4 (Fig. 3a, b) of the domain. Asparagine and tyrosine contact residues in the linker part of the C-terminal tail of Ran,



**Figure 3** Structure of RanBD1. **a**, Sequence alignment and secondary structure elements of Ran-binding domains and residues in RanBD1 from RanBP2 that are involved in contact with Ran-GppNHp, with the four principal contact regions (indicated by triangles in different colours) described in the text. Van der Waals contacts are indicated by black dots. **b**, Ribbon plot of RanBD1 (magenta) and of the PH domain from Bruton's tyrosine kinase (BTK, cyan) after superimposition to show the same orientation. The r.m.s. deviation is 1.4 Å for 100 C $\alpha$  atoms.  $\beta$ -Strand 5 of the normal PH domain is missing in RanBD1. The extra residues of BTK to the right of the  $\beta$ -barrel are inserted between  $\beta$ -strands 5 and 6 of BTK (as compared to RanBD1). The extra residues of BTK forming the BTK motif are omitted for clarity and are indicated by the dotted line extending from the C terminus of the BTK PH domain. The positions of InsP<sub>3</sub> (IP3) residues from the PH domain of rat phospholipase C $\delta$  with InsP<sub>3</sub> bound (PDB accession code, 1mai) and that of  $\beta$ -spectrin with InsP<sub>3</sub> bound (PDB accession code, 1btn) are depicted by yellow ellipses. Only part of the N terminus of RanBD1 is shown for clarity.

whereas cysteine, alanine, histidine and isoleucine pack tightly against other residues in the hydrophobic core of RanBD1, anchoring this region to the core to present a Ran-binding surface.

### The complex

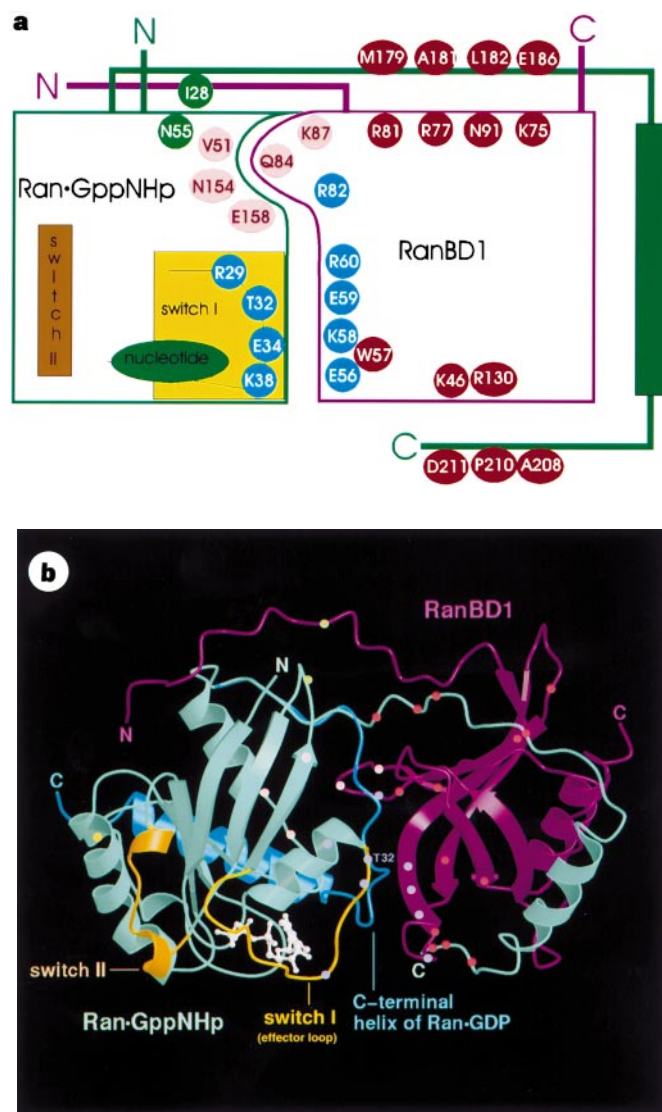
The complex between Ran-GppNHp and RanBD1, which has a dissociation constant of 4.3 nM (C. Villa-Braslavsky, unpublished results), buries 5,640 Å<sup>2</sup> of solvent-accessible surface area (calculated with a probe sphere of 1.7 Å using the Connolly algorithm)<sup>33</sup>, which is very large compared with that in other complexes with similar affinity. When Ran-GppNHp binds to RanBP1, there is a marked change in the circular dichroism spectrum<sup>34</sup>, indicating that complex formation causes a large conformational change in either Ran or RanBD1, or in both; however, as the structures of uncomplexed RanBD1 and Ran-GppNHp are unsolved, their contributions cannot be separated. Considering the change in the structure of Ran

between Ran-GDP and the RanBD1–Ran-GppNHp complex (Fig. 2b), at least part of the spectral change must be due to Ran. The contact between Ran and RanBD1 has four main contact sites (Fig. 4). First, the effector region of Ran around residues 29–38<sup>ran</sup> interacts with the conserved <sup>57</sup>WKER motif of RanBD1. When the structure of the PH domain of rat phospholipase Cδ bound to inositol 1,4,5-trisphosphate (InsP<sub>3</sub>) (ref. 35) is superimposed on the RanBD1 structure, InsP<sub>3</sub> is located near residue 32<sup>ran</sup> of Ran and the corresponding loops of the PH/BD1 domains are involved in binding (Fig. 3b). Likewise, the InsP<sub>3</sub>-binding site of β-spectrin<sup>36</sup> would also be in the vicinity of the effector loop (Fig. 3b; ref. 4). Second, the linker part of the C terminus of Ran (173–190<sup>ran</sup>) runs along the loop between β-strands 6 and 7 (Figs 3b, 4), and the helical part (191–206<sup>ran</sup>) lies in a groove on the surface of the RanBD1 core. Third, the N-terminal part of RanBD1 reaches across the Ran molecule and makes contact with the region around the β2–β3 loop of Ran and with residues 116<sup>ran</sup>/168<sup>ran</sup> (Fig. 4). Fourth, the loop 82–87 (<sup>80</sup>MRR loop) of RanBD1 extends from the core towards residues 154<sup>ran</sup>/158<sup>ran</sup> (Fig. 4). The long random-coil regions in Ran and in RanBD1 in the complex help to create an impression of a molecular ‘embrace’ between the two proteins as Ran wraps its C-terminal end around RanBD1 to form an almost complete turn.

In Ran-GDP, contacts between the C-terminal helix and the core of Ran are mainly hydrophobic contacts, as they are in the complex, where the helix is shorter by three residues, with the core of RanBD1. Potential hydrogen-bond interactions are in both cases in the linker region preceding the helix or in the regions without secondary structure that follow the helix. The accessible surface area buried by the helix in the complex (392 Å<sup>2</sup>) is about half that buried in Ran-GDP (800 Å<sup>2</sup>), supporting the idea that the C-terminal helix of Ran binds weakly to RanBD1, as we infer from the poor density for the Ran C terminus in the second molecule of the asymmetric unit. The N terminus of RanBD1 (residues 283–302), on the other hand, buries about 900 Å<sup>2</sup> in the complex, mostly by means of van der Waals contacts, so it probably contributes more to its stability than the C-terminal helix does. The DEDDDL motif in the Ran C terminus (211–216<sup>ran</sup>) and the N-terminal acidic sequence in RanBD1 (residues 10–14) both contribute significantly to the affinity of Ran for RanBD1: a deletion in the Ran C terminus<sup>24</sup> reduces the affinity of Ran-GppNHp for RanBP1 by ~8,000-fold (ref. 34); the effect of a deletion in the RanBD1 N-terminal region is shown (see below).

The core of the PH-like domain of RanBD1 binds to Ran around the effector region. It is thus sensitive to the nucleotide-bound state of Ran and may be able to trigger other, secondary binding events. Although the number of interactions between the core of the GTP-binding and RanBD domains is small, the interacting <sup>57</sup>WKER motif of RanBD1 is invariant, and the arginine residue in the <sup>80</sup>MRREQ motif is almost invariant. Residues of the <sup>57</sup>WKER motif participate in several electrostatic interactions with residues 29–38<sup>ran</sup> from the effector region. Arg 82 in <sup>80</sup>MRR is on a tight turn between the end of β3 and the start of the very short β-sheet 4; this turn forms a wedge that penetrates the interface. Residues <sup>82</sup>REQ on the tip of this wedge are involved in hydrophilic interaction with the effector region, with Asn 154<sup>ran</sup> and Glu 158<sup>ran</sup> on Ran, and with other parts of RanBD1. Although none of these residues apart from arginine is invariant, the side chains are always hydrophilic and able to maintain a similar network of contacts.

RanBPs stabilize the GTP form of Ran by inhibiting both the intrinsic (in the presence of Mg<sup>2+</sup>) and EDTA-stimulated release of GTP (refs 8, 24), thereby preventing uncontrolled nucleotide exchange. In the complex, RanBD1 does not block the nucleotide-binding site; this suggests that the binding of the conserved <sup>57</sup>WKER motif of RanBD1 with the effector loop of Ran could cause this effect. Fixation of the effector loop in the GTP-bound conformation by RanBD1 could trap the bound nucleotide and might explain why this conformation is stabilized. Effector proteins such as RafRBD



**Figure 4** The Ran-RanBD1 complex structure. **a**, **b**, Schematic (**a**) and ribbon (**b**) representation of the interaction between Ran-GppNHp and RanBD1. Superimposed in **b** is the C-terminal linker and helix from the Ran-GDP structure (cyan), highlighting the large C-terminal conformational switch. Residues from four contact regions between Ran and RanBD1 are indicated by dots and coloured as in the sequence alignment in Fig. 3a. The yellow dot on the left indicates the location of the basic patch in Ran (residues 139–142<sup>ran</sup>) that might act as the docking site for the DEDDDL motif in Ran-GDP and the (invisible) N terminus of RanBD1.



(where RBD is Ras-binding domain) also inhibit guanine-nucleotide release from Ras or Rap by an indirect effect (ref. 37), which can be deduced from the crystal structure of the Rap–RafRBD complex<sup>38</sup>. In both complexes, a tyrosine residue (Tyr 32 from Ras and Tyr 39 from Ran) shields the phosphates from solvent and the phenolic hydroxyl interacts with a  $\gamma$ -phosphate oxygen.

The N terminus of RanBD1 runs past residues 168–171<sup>ran</sup> in the loop between helix  $\alpha$ 5 and the C-terminal linker region, continues past residues 55<sup>ran</sup>/56<sup>ran</sup> in loop  $\beta$ 2– $\beta$ 3 and the Ran N terminus, and reaches down to residues 116<sup>ran</sup>/168<sup>ran</sup>. The conserved <sup>17</sup>HFEPPVPL region furnishes most of the contacts, with the invariant Phe 18 and two valines (at positions 21 and 22) making hydrophobic contacts with Ran; the conserved Glu 19 and Pro 20 residues point into the solvent, indicating that this region binds either to another protein or to other parts of RanBD1 when it is not complexed to Ran. Mutation of Glu 37 in HTF9a (corresponding to Glu 29 following the HFEPPV motif in RanBD1) reduces the affinity of Ran tenfold (ref. 28). The N terminus of RanBD1 is flexible in the unbound RanBD1 molecule, as demonstrated by proton NMR (H. R. Kalbitzer, unpublished results) and by trypsin cleavage of the first 34 residues and the protection by Ran·GTP against digestion (C. Villa-Braslavsky *et al.*, manuscript in preparation); trypsinized RanBD1 is unable to bind Ran. This result, and the conservation of key residues, underscores the importance of the N terminus for binding to Ran. The N-terminal end of the structure defined by electron density starts with residues <sup>17</sup>HFE (Fig. 3a) and so coincides with the Ran-binding domain as defined by the conserved sequence in RanBDs (Fig. 3a).

The residues preceding <sup>17</sup>HFEPPV are acidic in all RanBDs examined so far, although their location and sequence are not conserved. Judging from the position of the first structurally defined residues (<sup>17</sup>HFEPP), this acidic stretch should occupy the space close to the basic patch on Ran (residues 139–142<sup>ran</sup>), where the acidic C terminus of Ran (211–216<sup>ran</sup>) is bound in the GDP form. This proximity is unlikely to be accidental and may function to displace the Ran C terminus and/or to enhance affinity for Ran.

### The C-terminal switch of Ran

The conformation of the Ran C terminus in our complex is unlike that in Ran·GDP: the linker region and the C-terminal helix containing the DEDDDL motif both embrace the RanBD1 molecule, and the C-terminal linker follows quite a different path

towards RanBD, beginning with residue 173<sup>ran</sup> (Figs 2b, 4). The C-terminal helix moves from being close to residues 132–141<sup>ran</sup> in the GDP form, away from the G domain towards residues 96–99 of the RanBD1 domain. The DEDDDL motif, which is not visible in the electron density, moves close to and probably makes contact with a very basic patch on RanBD1 formed by residues 44, 46, 108, 130 and 132 of RanBD1 (Fig. 5). This fits nicely with the observation that the DEDDDL motif is necessary for tight binding to RanBP1 (refs 24, 34) as well as for the RanBDs of RanBP2 (J. Kuhlmann, unpublished data).

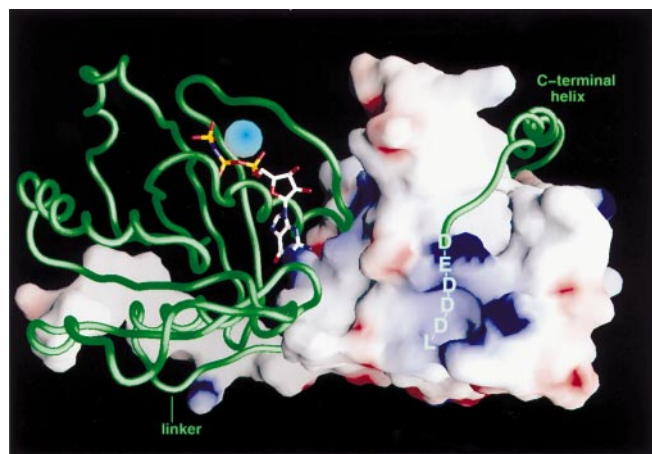
The conformational switch of the C-terminal helix triggered by the exchange of GDP for GTP on Ran increases the affinity for RanBD1 by  $\sim 10^4$ . The structure suggests that the force driving this molecular switch is the steric clash between the Ran linker region preceding the C-terminal helix and the effector loop following its rearrangement after Ran binds GTP, which principally involves residues 32–34<sup>ran</sup> from the effector loop and residues 182–184<sup>ran</sup> from the linker (Fig. 2c, d). If the linker maintains the position it occupies in the GDP-bound conformation, it would prevent the approach of the RanBD1 MRR motif that is involved in several interactions with Ran·GppNHp (Fig. 2c, d). In the GDP-bound form of Ran with the  $\beta$ -hairpin conformation of the effector loop, the linker positions the C-terminal helix in its energetically favoured position such that the C-terminal helix makes contacts with residues 124<sup>ran</sup>, 127<sup>ran</sup>, 133<sup>ran</sup>, 148<sup>ran</sup> and 155<sup>ran</sup>, and the DEDDDL motif is close to the basic patch on Ran<sup>17</sup>. In the GTP-bound conformation, the steric clash in the effector region could trigger a conformational change of the C terminus, the details of which will become apparent once the structure of uncomplexed Ran in the triphosphate conformation is solved.

The DEDDDL motif must be important for the C-terminal switch as it stabilizes the GDP-bound and destabilizes the GTP-bound forms. In the GTP-bound form, this motif seems to force the linker and C-terminal helix into an unfavourable association<sup>24</sup>. These unfavourable interactions are relieved by removing the DEDDDL motif or by binding of RanBDs, which also have a basic region designed for binding DEDDDL (Fig. 5). The switch mechanism used by Ran is comparable to the GTP–myristoyl switch used by Arf (ref. 23), in which a conformational change triggered by swapping the bound nucleotide blocks the binding site for the N-terminal myristoylated helix in Arf·GDP. This binding site is predominantly hydrophobic, like the C-terminal helix binding sites in Ran or RanBD1. In the case of Arf, the displacement of the helix favours the membrane interaction; in the case of Ran, it facilitates binding to other regulatory molecules. The structural studies on Arf and Ran show that introducing extra elements into the G-domain structure can generate a variety of modifications in the canonical switch mechanism used by Ras-related proteins.

### Implications for nucleocytoplasmic transport

Ran is unique among the Ras-related GTP-binding proteins in that it has a long extension at the C terminus of the G domain that ends in the conserved DEDDDL motif: this motif is necessary for transport processes across the nuclear pore<sup>39</sup> and it inhibits the binding of importin- $\beta$  to Ran, as its deletion increases the association rate constant of the interaction (C. Villa-Braslavsky *et al.*, manuscript in preparation). Further, RanBP1 increases ranGAP-mediated hydrolysis of GTP on Ran and the deletion of the DEDDDL motif abolishes this stimulation; we have shown that RanBD1 and the other RanBDs from RanBP2 behave similarly (C. Villa-Braslavsky *et al.*, manuscript in preparation).

The structure of the complex indicates that the main function of Ran-binding domains is probably to sequester the C-terminal extension of Ran (Fig. 5). We can assume that in the Ran·GTP conformation, the C-terminal acidic region hinders access by ranGAP to the active site and by binding to RanBDs it increases the affinity between ranGAP and Ran·GTP. By comparing the



**Figure 5** Molecular embrace and the DEDDDL motif. Surface representation of RanBD1, showing the basic region where the DEDDDL motif of Ran is expected to bind after the C terminus wraps itself around RanBD1. Ran is shown as a backbone (green), and GppNHp and the magnesium ion as ball and stick. The figure was produced using GRASP<sup>50</sup>.



Ran-RanBD1 complex with the structure of the Ras-RasGAP complex<sup>40</sup> and assuming a similar mechanism, the active site of Ran-GppNHP in the RanBD1 complex seems to be open enough for ranGAP to approach in the same way and participate in the phosphoryl-transfer reaction. ranGAP thus stimulates GTPase activity while the RanBDs presumably alter the  $K_m$  but not the  $k_{cat}$  of the reaction.

In the absence of a structural model for Ran-transport-factor complexes, the function of RanBPs as a 'dissociator' of nuclear-export complexes is not easy to explain. RanBP1 is necessary for the release of Ran-GTP from trimeric Ran-GTP-exportin-cargo complexes such as the Cas-importin complex<sup>6,9</sup>. Carboxypeptidase A digests of complexes of Ran-GppNHP and importin- $\beta$  show a limited digestion of the Ran C terminus up to Ala181<sup>ran</sup> (...FVAMPA<sup>181</sup>), which removes the C-terminal helix plus the linker region and shows that the C terminus is accessible in the complex; Ran-GTP on its own is not digested. A monoclonal antibody against the C-terminal DEDDDL motif recognizes the antigen only in a Ran-GTP-importin complex and not in free Ran-GTP (Y. Yoneda, personal communication), indicating that the C-terminal end is not or only loosely connected to the G domain in the complex. Heterotrimeric complexes can form readily between Ran, RanBP and importin- $\beta$ -like proteins<sup>41</sup>. As RanBDs bind strongly to the C terminus, the accessible C terminus of Ran-exportin complexes should be recognized by a RanBD which then unravels the complex by wrapping the complete C-terminal end around itself, loosening the connection with the exportin-cargo complex. As importin- $\beta$ -like proteins block the ranGAP-mediated GTPase reaction on Ran (refs 5, 7), importin and ranGAP binding to Ran could be mutually exclusive. Dissociation of the Ran-GTP-transport factor complex under the influence of RanBD1 would prime the resulting Ran-GTP-RanBD complex for switching off by ranGAP. As RanBP2 contains four Ran-binding domains and is covalently modified by ranGAP through the small protein SUMO, the export reaction is terminated at the exit of the nuclear pore. Cytoplasmic RanBP1 and ranGAP would then act as the last line of defence against escaping Ran-cargo complexes. It may be that the developing Ran-GDP complexes stay at the nuclear entrance to be assembled into import complexes<sup>16</sup>, given that ternary RanGDP-RanBD-importin complexes have sufficient affinity<sup>41</sup>.

We have shown that Ran is a new type of molecular switch in which the canonical mechanism involving the effector loop is coupled to a large conformational change of the long C-terminal extension, which includes the crucial acidic DEDDDL motif, exposing it to RanBDs at the nuclear pore exit. RanBDs wrap the C-terminal end of Ran around themselves, causing Ran-transport factor complexes to dissociate and ranGAP to hydrolyse GTP. We conclude that RanBDs are a new type of effector protein to interact with GTP-binding proteins that act as terminators of Ran-effector interaction. □

## Methods

**Protein preparation, data collection and reduction.** Ran protein was expressed and purified as described<sup>34</sup>. A fusion protein of glutathione-S-transferase (GST) with RanBP2<sup>RanBD1</sup> was expressed from *E. coli* strain BL21, induced with 100  $\mu$ M IPTG overnight at 30 °C, lysed by sonication and purified over a glutathione-Sepharose column. Exchange of bound GDP with the non-hydrolysable GTP analogue GppNHP (5'-guanosyl-[ $\beta$ , $\gamma$ -imidol]-triphosphate) on Ran was achieved by incubating Ran-GDP with a twofold excess of GppNHP and 2.5 units of alkaline phosphatase per mg protein overnight at 4 °C in 20 mM potassium phosphate buffer, pH 7.4, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 mM ZnCl<sub>2</sub>, 1 mM dithioerythritol (DTE). Because of the limited yield of Ran-GppNHP (~70%), complexes were formed on a glutathione GSH-Sepharose column with prebound GST-RanBP2<sup>RanBD1</sup> and a twofold excess of Ran-GppNHP over RanBD1 to ensure complete saturation of the binding domain with Ran-GppNHP. The GST-RanBD1-Ran-GppNHP complex was

washed with binding buffer (20 mM potassium phosphate, pH 7.4, 5 mM MgCl<sub>2</sub>, 1 mM DTE) and cleaved with 20 units thrombin per mg GST overnight at 4 °C. The RanBD1-Ran-GppNHP complex was eluted with binding buffer.

Crystals were grown from 18% PEG 4000, 250 mM ammonium sulphate, 100 mM MES buffer, pH 6.25. The cryosolution consisted of 18% PEG 4000, 100 mM MES, pH 6.25, 250 mM ammonium acetate and 10% PEG 1000. Crystals with two complexes in the asymmetric unit, space group C2 and unit-cell dimensions of  $a = 121.3$ ,  $b = 139.5$ ,  $c = 91.7$ ,  $\beta = 135.8$  were diffracting to at least 2.9 Å at the BW7B beamline at DESY, Hamburg. Data were collected at 100 K and a wavelength of 0.8342 Å with a 34.5 cm MAR scanner (18-cm plate) and processed with XDS (ref. 42). To improve the completeness of the data set at low resolution, the synchrotron data set was merged with in-house data collected on a rotating-anode source equipped with a Siemens HiStar area detector. The  $R_{merge}$  was slightly higher after merging (9.1 compared to 5.7% for the synchrotron data by itself), but the more complete data showed a better convergence in refinement, improved maps and led to a lower  $R_{free}$ . The merged data set is 86% complete to 2.9 Å, with a redundancy of 4.5; the outer shell (3.0–2.9 Å) is 66% complete with an  $I/\sigma$  of 3.1.

**Molecular replacement.** The two Ran molecules in the asymmetric unit were found using AMORE of the CCP4 suite of crystallography programs<sup>43</sup> and Ran-GDP (ref. 17) as a model, with the effector loop and C terminus removed. The correlation for the correct solutions was 30.9 with a corresponding R-factor of 47.6%. Maps calculated with omitted nucleotides showed clear density for the two GppNHP molecules and two magnesium ions. The known part of the structure (Ran) represents ~55% of the total mass of the Ran-RanBD1 complex.

**Density modification and refinement.** Subsequent averaging and solvent flattening with DM (ref. 43) allowed building of the missing regions of Ran and the tracing of the RanBD1 moieties with program O (ref. 44). NCS (non-crystallographic symmetry) restraints (except for residues 17–29 of RanBD1 and residues 186–211 of Ran) were applied throughout the refinement of the structure using CNS (ref. 45). Restrained temperature factors were refined while checking the free R-factor for convergence. The current R-factor for 20,286 reflections is 25.2%; the free R-factor is 30.4%, with 10% reflections being excluded from the refinement for cross-validation. Only very weak electron density was visible for region 188–216 of Ran in molecule 2, which was therefore omitted from the model.

Received 6 November 1998; accepted 1 February 1999.

- Görlich, D. Transport into and out of the cell nucleus. *EMBO J.* **17**, 2721–2727 (1998).
- Hartmann, E. & Görlich, D. A Ran-binding motif found in nuclear pore proteins. *Trends Cell Biol.* **5**, 192–193 (1995).
- Dingwall, C., Kanders-Lewis, S. & Séraphin, B. A family of Ran binding proteins that includes nucleoporins. *Proc. Natl Acad. Sci. USA* **92**, 7525–7529 (1995).
- Rexach, M. & Blobel, G. Protein import into nuclei—association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell* **83**, 683–692 (1995).
- Görlich, D., Pante, N., Kutay, U., Aebi, U. & Bischoff, F. R. Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.* **15**, 5584–5594 (1996).
- Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, J. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
- Floor, M. & Blobel, G. The nucleic transport factor karyopherin-beta binds stoichiometrically to Ran-GTP and inhibits the Ran GTPase-activating protein. *J. Biol. Chem.* **271**, 5313–5316 (1996).
- Bischoff, F. R., Krebber, H., Smirnova, E., Dong, W. & Pöschel, H. Co-activation of RanGTPase and inhibition of GTP dissociation by Ran-GTP binding protein RanBP1. *EMBO J.* **14**, 705–715 (1995).
- Bischoff, F. R. & Görlich, D. RanBP1 is crucial for the release of Ran-GTP from importin- $\beta$ -related nuclear transport factors. *FEBS Lett.* **419**, 249–254 (1997).
- Wilken, N., Senécal, J.-L., Scheer, U. & Dabauvalle, M.-C. Localization of the Ran-GTP binding protein RanBP2 at the cytoplasmic side of the nuclear pore complex. *Eur. J. Cell Biol.* **68**, 211–219 (1995).
- Delphin, C., Guan, T., Melchior, F. & Gerace, L. Ran-GTP targets p97 to RanBP2, a filamentous protein localized at the cytoplasmic periphery of the nuclear pore complex. *Mol. Biol. Cell* **8**, 2379–2390 (1997).
- Yokoyama, N. *et al.* A giant nucleoporin protein that binds Ran/TC4. *Nature* **376**, 184–188 (1995).
- Wu, J., Matunis, M. J., Kraemer, D., Blobel, G. & Coutavas, E. Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J. Biol. Chem.* **270**, 14209–14213 (1995).
- Radu, A., Moore, M. S. & Blobel, G. The peptide repeat domain of nucleoporin Nup98 functions as docking site in transport across the nuclear pore complex. *Cell* **81**, 215–222 (1995).
- Melchior, F., Guan, T., Yokoyama, N., Nishimoto, T. & Gerace, L. GTP hydrolysis by Ran occurs at the nuclear pore complex in an early step of protein import. *J. Cell Biol.* **131**, 571–581 (1995).
- Melchior, F. & Gerace, L. Two-way trafficking with Ran. *Trends Cell Biol.* **8**, 175–179 (1998).
- Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. Crystal structure of human Ran complexed with GDP. *Nature* **374**, 378–381 (1995).
- Amor, J. C. *et al.* Structure of the human ADP-ribosylation factor I complexed with GDP. *Nature* **372**, 704–708 (1994).
- Greasley, S. E. *et al.* The structure of rat ADP-ribosylation factor-1 (ARF-1) complexed to GDP determined from two different crystal forms. *Nature Struct. Biol.* **2**, 797–806 (1995).
- Abel, K., Yoder, M. D., Hilgenfeld, R. & Jurnak, F. An alpha to beta conformational switch in EF-TU. *Structure* **4**, 1153–1159 (1996).

21. Polekhina, G. *et al.* Helix unwinding in the effector of elongation factor EF-TU-GDP. *Structure* **4**, 1141–1151 (1996).
22. Wittinghofer, A. & Pai, E. F. The structure of Ras protein: a model for a universal molecular switch. *Trends Biochem. Sci.* **16**, 382–387 (1991).
23. Goldberg, J. Structural basis for activation of ARF GTPase: Mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell* **95**, 237–248 (1998).
24. Richards, S. A., Lounsbury, K. M. & Macara, I. The C terminus of the nuclear RAN/TC4 GTPase stabilizes the GDP-bound state and mediates interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. *J. Biol. Chem.* **270**, 14405–14411 (1995).
25. Klebe, C., Bischoff, F. R., Ponstingl, H. & Wittinghofer, A. Interaction of the nuclear GTP-binding protein Ran with its regulatory proteins RCC1 and RanGAP1. *Biochemistry* **34**, 639–647 (1995).
26. Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M. & Görlich, D. NTF2 mediates nuclear import of Ran. *EMBO J.* **17**, 6587–6598 (1998).
27. Stewart, M., Kent, H. M. & McCoy, A. J. Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and GTP-bound form of the Ras family GTPase Ran. *J. Mol. Biol.* **277**, 535–646 (1998).
28. Beddow, A. L., Richards, S. A., Orem, N. R. & Macara, I. G. The Ran/TC4 GTPase-binding domain: Identification by expression cloning and characterization of a conserved sequence motif. *Proc. Natl Acad. Sci. USA* **92**, 3328–3332 (1995).
29. Saraste, M. & Hyvönen, M. Pleckstrin homology domains: a fact file. *Curr. Opin. Struct. Biol.* **5**, 403–408 (1995).
30. Hyvönen, M. & Saraste, M. Structure of the PH domain and BTK motif from Bruton's tyrosine kinase: Molecular explanations for X-linked agammaglobulinaemia. *EMBO J.* **16**, 3396–3404 (1997).
31. Touhara, K., Inglese, J., Pitcher, J. A., Shaw, G. & Lefkowitz, R. J. Binding of G protein  $\beta\gamma$ -subunits to pleckstrin homology domains. *J. Biol. Chem.* **269**, 10217–10220 (1994).
32. Schlenstedt, G., Wong, D. H., Koeppe, D. M. & Silver, P. A. Mutants in a yeast Ran binding protein are defective in nuclear transport. *EMBO J.* **14**, 5367–5378 (1995).
33. Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16**, 548–558 (1983).
34. Kuhlmann, J., Macara, I. & Wittinghofer, A. Dynamic and equilibrium studies on the interaction of Ran with its effector RanBP1. *Biochemistry* **36**, 12027–12035 (1997).
35. Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. Structure of the high affinity complex of inositol triphosphate with a phospholipase C pleckstrin homology domain. *Cell* **83**, 1037–1046 (1995).
36. Hyvönen, M. *et al.* Structure of the binding site for inositol phosphates in a PH domain. *EMBO J.* **14**, 4676–4685 (1995).
37. Herrmann, C., Martin, G. A. & Wittinghofer, A. Quantitative analysis of the complex between p21<sup>ras</sup> and the Ras-binding domain of the human Raf-1 protein kinase. *J. Biol. Chem.* **270**, 2901–2905 (1995).
38. Nassar, N. *et al.* The 2.2 Å crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature* **375**, 554–560 (1995).
39. Lounsbury, K. M., Richards, S. A., Carey, K. L. & Macara, I. G. Mutations within the Ran/TC4 GTPase-effects on regulatory factor interactions and subcellular localization. *J. Biol. Chem.* **271**, 32834–32841 (1996).
40. Scheffzek, K. *et al.* The Ras–RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1998).
41. Chi, N. C., Adam, E. J. H., Visser, G. D. & Adam, S. A. RanBP1 stabilizes the interaction of Ran with p97 in nuclear protein import. *J. Cell. Biol.* **135**, 559–569 (1996).
42. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell contents. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
43. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
45. Brunger, A. T. *et al.* Crystallography and NMR system: A new software system for macromolecular structure determination. *Acta Crystallogr. C* **54**, 905–921 (1998).
46. Esnouf, R. M. An extensively modified version of MOLSCRIPT that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 132–134 (1997).
47. Pai, E. F. *et al.* Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *EMBO J.* **9**, 2351–2359 (1990).
48. Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
49. Merritt, E. A. & Murphy, M. E. P. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
50. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

**Acknowledgements.** We thank A. Scherer for help with the original crystallization, R. Schebaum for secretarial assistance, and the staff of the EMBL outstation in Hamburg for help with the data collection. This study was supported by HFSP and the EG (A.W.).

Correspondence and requests for materials should be addressed to A.W. (e-mail: alfred.wittinghofer@mpi-dortmund.mpg.de). The atomic coordinate has been deposited at the Brookhaven Protein Databank under code 1rrp.

# Superconductivity mediated by spin fluctuations in the heavy-fermion compound UPd<sub>2</sub>Al<sub>3</sub>

M. Jourdan, M. Huth & H. Adrian

Institute of Physics, Johannes Gutenberg-University,  
Staudingerweg 7, 55099 Mainz, Germany

It is well known that any weak attractive electron–electron interaction in metals can in principle cause the formation of Cooper pairs, which then condense into a superconducting ground state<sup>1</sup>. In conventional superconductors, this attractive interaction is mediated by lattice vibrations (phonons). But for the heavy-fermion and high-temperature superconductors, alternative pairing interactions are considered to be possible<sup>2</sup>. For example, the low-temperature properties of heavy-fermion systems are dominated by antiferromagnetic spin fluctuations, which have been considered theoretically<sup>3</sup> as a possible cause for Cooper-pair formation. This picture recently received some experimental support: the resistivity behaviour under pressure of two cerium-based heavy-fermion compounds was shown to be consistent with a magnetically mediated pairing mechanism<sup>4</sup>. Here we use tunnelling spectroscopy to investigate the superconducting order parameter of a uranium-based heavy-fermion superconductor—epitaxial thin films of UPd<sub>2</sub>Al<sub>3</sub>. Our observation of a strong-coupling feature in the tunnelling conductivity, combined with recent inelastic neutron scattering data<sup>13–15</sup> strongly suggest a pairing interaction mediated by spin fluctuations.

Heavy-fermion systems are intermetallic compounds with strong *f*-electron correlations resulting in a pronounced increase of the effective charge-carrier mass to about 100 free electron masses. Due to these strong *f*-electron correlations, the realization of pairing states with nodes at the pair centre were considered likely for energetic reasons<sup>5</sup>. This might be seen as an analogy to the combined hard-core repulsion spin-fluctuation-induced superfluidity of <sup>3</sup>He. Whereas the *p*-wave superfluidity in <sup>3</sup>He is mediated by ferromagnetic spin fluctuations, the interaction in heavy-fermion superconductors that we discuss here is of antiferromagnetic origin.

Attempts to investigate the superconducting order parameter of heavy-fermion systems by spectroscopic means suffered from degraded superconductivity in the surface region of the examined bulk samples<sup>6,7</sup>. Now that epitaxial heavy-fermion thin films can be

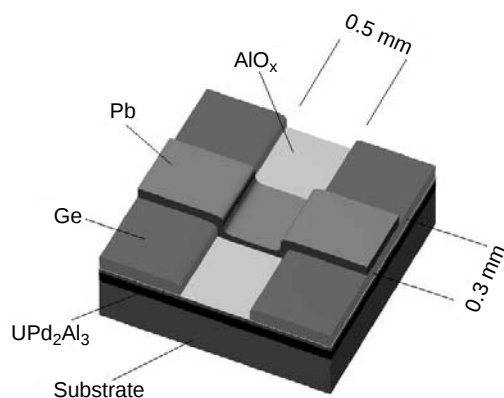
grown, it is possible to prepare planar superconductor–insulator–normal conductor tunnelling junctions entirely in vacuum; this produces clean interfaces. Such interfaces proved to be essential owing to the short superconducting coherence length of these compounds (about 10 nm). For such short coherence lengths any kind of surface degradation severely interferes with the tunnelling process.

We investigated UPd<sub>2</sub>Al<sub>3</sub>, a heavy-fermion system in which superconductivity below the transition temperature *T*<sub>c</sub> = 2 K coexists with antiferromagnetic order below the Néel temperature *T*<sub>N</sub> = 14.5 K (ref. 8). Epitaxial thin films of UPd<sub>2</sub>Al<sub>3</sub> were grown in (0001) orientation on single-crystalline LaAlO<sub>3</sub> (111) substrates by molecular beam epitaxy as described elsewhere<sup>9</sup>. The films, which gave good tunnelling junctions, showed a slightly reduced *T*<sub>c</sub> of 1.6 K. As compared to bulk samples, antiferromagnetism in the films is also observed at a reduced *T*<sub>N</sub> of 12 K (ref. 10).

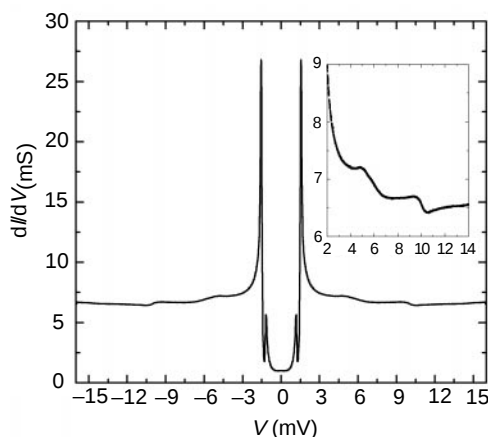
Tunnelling junctions were obtained by sputtering a layer of aluminium, about 4 nm thick, on top of the UPd<sub>2</sub>Al<sub>3</sub> film. The Al was then oxidized in an oxygen glow discharge, resulting in an insulating tunnelling barrier. A four-terminal cross-junction geometry was realized by covering almost the whole film (but leaving a stripe uncovered) with an insulating layer, ~200 nm thick, of amorphous Ge, employing a shadow mask technique: this step defined the base electrode. Finally, Pb was evaporated through a shadow mask giving the counter-electrode. A diagram of the junction geometry is shown in Fig. 1. Typical junction areas are 0.5 × 0.3 mm.

The bias-voltage-dependent differential conductivity of a superconductor–insulator–normal conductor tunnelling junction is, in the simplest case, a direct measure of the thermally smeared density of states of the investigated superconductor. In the case of a superconductor–insulator–superconductor tunnelling junction, a folding of the energy-dependent density of states of both electrodes is observed. Strong-coupling superconductivity manifests itself by the appearance of additional structures in the differential conductivity outside the energy gap region. In conventional superconductors, these structures are due to the coupling of the charge carriers to phonon modes.

In Fig. 2 the differential conductivity of a UPd<sub>2</sub>Al<sub>3</sub>–AlO<sub>x</sub>–Pb junction is shown, measured by a standard a.c.-modulation technique in a <sup>3</sup>He cryostat. The characteristic results for a superconductor–insulator–superconductor tunnelling junction are obtained, showing conductivity peaks related to the sum ( $\Delta_1 + \Delta_2$ ) and the difference ( $\Delta_1 - \Delta_2$ ) of the energy gaps of both electrodes. Additionally, the phononic strong-coupling modulations typical for Pb are shown in the inset. We note that by using Pb as a counter-electrode, it is possible to demonstrate that the junctions are in the tunnelling regime; furthermore, the observation



**Figure 1** Diagram of a UPd<sub>2</sub>Al<sub>3</sub>–AlO<sub>x</sub>–Pb tunnel junction. Dimensions are not to scale.



**Figure 2** Differential conductivity of a UPd<sub>2</sub>Al<sub>3</sub>–AlO<sub>x</sub>–Pb tunnel junction at *T* = 0.3 K. Inset, phononic strong coupling structures of Pb (same axis parameters as main plot).



of the well-known tunnelling features of Pb reflects the high quality of the contact.

Applying a magnetic field which is overcritical for Pb, but which is well below the upper critical field  $H_{c2}$  of  $\text{UPd}_2\text{Al}_3$ , a superconductor–insulator–normal conductor junction is obtained. This allows us to investigate directly the tunnelling density of states of the heavy-fermion superconductor. Figure 3 shows the differential conductivity of a  $\text{UPd}_2\text{Al}_3$ – $\text{AlO}_x$ –Pb tunnelling junction at various temperatures in an applied constant magnetic field of  $\mu_0 H = 0.3 \text{ T}$ . The observed energy gap closes with increasing temperature as well as with increasing magnetic field at  $T_c = 1.6 \text{ K}$  and  $\mu_0 H_c = 3.5 \text{ T}$ , respectively. This is in correspondence with the resistively measured superconducting transition of the  $\text{UPd}_2\text{Al}_3$  film. We also observe a reduced conductivity followed by a shallow maximum at bias voltage  $V \approx 1.22 \text{ meV}$ . This conductivity modulation is reduced with increasing temperature or magnetic field. It disappears in conjunction with the conductivity maximum at about  $V = \Delta$ . We assume that a shallow dip in the junction conductivity in the normal-conducting state, which disappears close to  $T_N$ , is caused by a reduction of the normal density of states due to a coupling to the spin-fluctuation spectrum.

In order to infer the size of the energy gap  $\Delta$  from our measurements, the junction conductivity  $\sigma$  (normalized to the conductivity at temperature  $T = 1.8 \text{ K} > T_c$ ) can be fitted successfully by the Dynes formula<sup>11</sup>:  $\sigma(V, \Gamma) = \text{Re}\{(V - i\Gamma)/[(V - i\Gamma)^2 - \Delta^2]^{1/2}\}$  with a broadening parameter  $\Gamma$ . The normalized conductivity in the superconducting state at  $T = 0.3 \text{ K}$  and the corresponding Dynes fit are shown in Fig. 4. We obtain an extrapolated gap width of  $\Delta_0 = 235 \mu\text{eV}$  at  $T = 0 \text{ K}$  and a broadening parameter of  $\Gamma = 35 \mu\text{eV}$ . The broadening parameter is independent of temperature and modulation voltage  $V_{ac} < 20 \mu\text{V}$ . The introduction of  $\Gamma$  is generally motivated by the assumption of quasiparticle lifetime effects. Another possible reason for a temperature-independent broadening parameter in an anisotropic superconductor is given by contributions of a range of tunnelling directions, resulting in an angular distribution of  $\Delta$  values. However, this can be excluded in the present case, as the tunnelling barrier, with an area resistance of  $\rho_s = 23 \Omega \text{ mm}^2$ , is well defined and the film surface is reasonably flat. The tunnelling current is therefore dominated by contributions along the crystallographic  $c$ -axis of  $\text{UPd}_2\text{Al}_3$ .

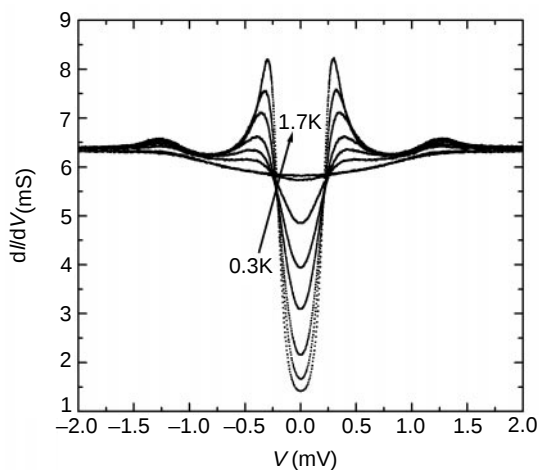
The most important experimental observation is the conductivity modulation at  $V \approx 1.22 \text{ meV}$  which cannot be described by the simple Dynes treatment as shown in the inset of Fig. 4. But it can be understood in a straightforward manner by assuming a strong-coupling effect between the charge carriers and an antiferromagnetic spin fluctuation. Conventional coupling to a phonon

mode can be ruled out, as the observed feature appears at an energy far below typical phonon energies. These might be estimated by the Debye temperature,  $\sim 150 \text{ K}$ , of  $\text{UPd}_2\text{Al}_3$  as determined by specific-heat measurements<sup>12</sup>. No structures in the differential conductivity in the corresponding energy region were observed. Spin fluctuations, on the other hand, should be considered, as inelastic neutron scattering experiments have found a gapped dispersive spin excitation that amounts to  $\Delta E = 1.5 \text{ meV}$  at the magnetic Bragg-point  $Q = (0, 0, 1/2)$  (refs 13–15). Although this is not exactly the energy at which the maximum in the tunnelling conductivity appears, this spin excitation probably represents the mode to which the superconducting order parameter couples.

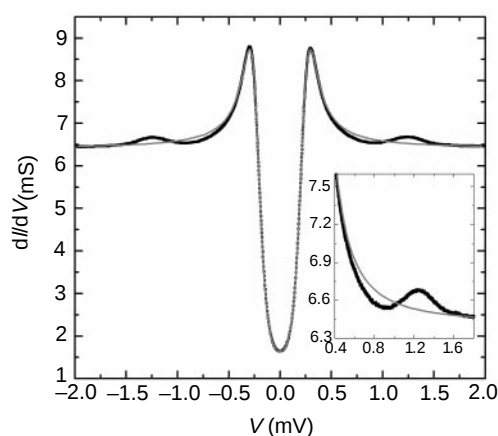
Assuming a coupling of the superconducting order parameter to this spin excitation, our experimental findings can qualitatively be explained in the framework of the Eliashberg theory of strong-coupling superconductivity<sup>16</sup>. Assuming a single peak in the spin-fluctuation density of states at  $\omega_0$  an increased pairing interaction results which is then reflected in an increased tunnelling density of states at this energy. In the case of spin-fluctuation-mediated  $d$ -wave superconductivity, a critical energy  $\omega_c$  is introduced, below which the spin fluctuations are pair-breaking<sup>17</sup>. This is in contrast to a phononic coupling mechanism where at excitation energies  $\omega$  less than  $\omega_0$  a net attractive interaction is created. With a critical energy  $\omega_c$  of the order of  $\omega_0$ , this special spin-wave pair-breaking effect causes the reduction of the tunnelling conductivity to below the Dynes fit (see Fig. 4 inset). At energies  $\omega > \omega_c$ , the attractive electron–electron interaction increases, resulting in an increase of the tunnelling density of states.

The present tunnelling results alone cannot prove that the pairing mechanism is solely due to spin fluctuations. Further experimental investigations need to be performed; specifically, detailed studies of the spin-fluctuation and phonon spectrum of  $\text{UPd}_2\text{Al}_3$ . But the absence of phononic structures strongly suggests spin-fluctuation pairing as the dominant mechanism. It is commonly believed that antiferromagnetic spin fluctuations suppress conventional isotropic pairing but promote anisotropic singlet pairing, resulting in nodes of the energy gap<sup>3</sup>. This is an agreement with measurements of the NMR relaxation time<sup>18,19</sup> and thermal conductivity<sup>20</sup> which infer an energy gap that vanishes along lines on the Fermi surface. From our direction-sensitive measurement of the energy gap along the crystallographic  $c$ -axis of  $\text{UPd}_2\text{Al}_3$ , we can draw further conclusions concerning the symmetry of the order parameter.

In general, the order parameter can be classified according to its symmetry by arguments based on group theory<sup>21</sup>. Considering all the possible even-parity  $d$ -wave irreducible representations of the hexagonal point group  $D_{6h}$  to which  $\text{UPd}_2\text{Al}_3$  belongs, only one



**Figure 3** Differential conductivity of a  $\text{UPd}_2\text{Al}_3$ – $\text{AlO}_x$ –Pb tunnel junction at various temperatures. (Temperatures  $T = 0.3 \text{ K}, 0.5 \text{ K}, 0.7 \text{ K}, \dots, 1.7 \text{ K}$ .) Superconductivity of Pb is suppressed in a magnetic field of  $\mu_0 H = 0.3 \text{ T}$ .



**Figure 4** Normalized differential conductivity of a  $\text{UPd}_2\text{Al}_3$ – $\text{AlO}_x$ –Pb tunnel junction. Black trace, experimental results for  $T = 0.3 \text{ K}$ ,  $B = 0.3 \text{ T}$ ; grey trace, fit to the Dynes formula with  $\Delta = 235 \mu\text{eV}$ ,  $\Gamma = 35 \mu\text{eV}$  (see text). Inset, spin-fluctuation strong-coupling structures of  $\text{UPd}_2\text{Al}_3$  (same axis parameters as main plot).

exists that does not show a gap node along the  $c$ -axis. We deduce then that only the fully symmetric  $A_{1g}$  representation can be realized. This representation includes the isotropic  $s$ -wave and higher-order basis functions ( $k_x^2 + k_y^2$ ) and  $k_z^2$ , from which a  $d$ -wave order parameter with a line node around the  $c$ -axis can be constructed. Order parameters of odd parity are to be ruled out, owing to the strong Pauli limiting that is observed in  $UPd_2Al_3$  (ref. 22).

Our experiments therefore suggest that the superconductivity in the heavy-fermion system  $UPd_2Al_3$  is mediated by antiferromagnetic spin fluctuations inducing a presumably  $d$ -wave order parameter without any unconventional symmetry reduction. □

Received 14 September; accepted 14 December 1998.

1. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).
2. Cox, D. L. & Maple, M. B. Electronic pairing in exotic superconductors. *Phys. Today* **275**, 32–40 (1995).
3. Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation-mediated even-parity pairing in heavy-fermion superconductors. *Phys. Rev. B* **34**, 6554–6556 (1986).
4. Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
5. Grewe, N. & Steglich, F. in *Handbook on the Physics and Chemistry of Rare Earth* (eds Gscheidner, K. A. & Eyring, L.) Vol. 14, Ch. 97 (Elsevier Science, Amsterdam, 1991).
6. Gloos, K. *et al.* Scaling behaviour of point contacts between a tungsten tip and the heavy-fermion superconductors. *J. Low. Temp. Phys.* **105**, 37–65 (1996).
7. von Loehneysen, H. Probing the energy gap of the heavy-fermion superconductors. *Physica B* **218**, 148–156 (1996).
8. Geibel, C. *et al.* Heavy-fermion superconductivity at  $T_c = 2$  K in the antiferromagnet  $UPd_2Al_3$ . *Z. Phys. B* **84**, 1–2 (1991).
9. Huth, M., Kaldowski, A., Hessert, J., Heske, C. & Adrian, H.  $UPd_2Al_3$  heavy fermion superconducting films. *Physica B* **199** & **200**, 116–118 (1994).
10. Hiess, A. *et al.* Magnetic properties of  $UPd_2Al_3$  thin films investigated by resonant magnetic x-ray scattering. *Physica B* (in the press).
11. Dynes, R. C., Narayanamurti, V. & Garno, J. Direct measurement of quasiparticle-lifetime broadening in a strong-coupled superconductor. *Phys. Rev. Lett.* **41**, 1509–1512 (1978).
12. Caspary, R. *et al.* Unusual ground state properties of  $UPd_2Al_3$ : Implications for the co-existence of heavy-fermion-superconductivity and local-moment antiferromagnetism. *Phys. Rev. Lett.* **71**, 2146–2149 (1993).
13. Sato, N. *et al.* Spin fluctuations in the heavy fermion superconductor  $UPd_2Al_3$  studied by neutron inelastic scattering. *J. Phys. Soc. Jpn* **66**, 1884–1887 (1997).
14. Metoki, N., Haga, Y., Koike, Y. & Onuki, Y. Superconducting energy gap observed in the magnetic excitation spectra of a heavy fermion superconductor  $UPd_2Al_3$ . *Phys. Rev. Lett.* **80**, 5417–5420 (1998).
15. Bernhoeft, N. *et al.* Magnetic fluctuations above and below  $T_c$  in the heavy fermion superconductor  $UPd_2Al_3$ . *Physica B* (in the press).
16. Parks, R. D. in *Superconductivity* Ch. 10 (Marcel Dekker, New York, 1969).
17. Millis, A. J., Sachdev, S. & Varma, C. M. Inelastic scattering and pair breaking in anisotropic and isotropic superconductors. *Phys. Rev. B* **37**, 4975–4986 (1988).
18. Tou, H. *et al.*  $d$ -wave superconductivity in antiferromagnetic heavy-fermion compound  $UPd_2Al_3$ —evidence from  $^{27}Al$  NMR/NQR studies. *J. Phys. Soc. Jpn* **64**, 725–729 (1995).
19. Matsuda, K., Kohori, Y. & Kohara, T. Observation of  $^{105}Pd$  NMR and NQR signals in the heavy-fermion superconductor  $UPd_2Al_3$ . *Phys. Rev. B* **55**, 15223–15227 (1997).
20. Hiroi, M. *et al.* Thermal conductivity of a heavy fermion superconductor  $UPd_2Al_3$  single crystal. *J. Phys. Soc. Jpn* **66**, 1595–1598 (1997).
21. Yip, S. & Garg, A. Superconducting states of reduced symmetry: General order parameters and physical implications. *Phys. Rev. B* **48**, 3304–3308 (1993).
22. Hessert, J. *et al.* Temperature and angular dependence of the upper critical field of  $UPd_2Al_3$  thin films. *Physica B* **230–232**, 373–376 (1997).

**Acknowledgements.** We thank P. Haibach for experimental support. This work was supported by the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to M.H. (e-mail: huth@mail.uni-mainz.de).

## Coherent displacement of atoms during ion irradiation

K. Nordlund\*, J. Keinonen\*, M. Ghaly† & R. S. Averback†

\* Accelerator Laboratory, PO Box 43, FIN-00014 University of Helsinki, Finland

† Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

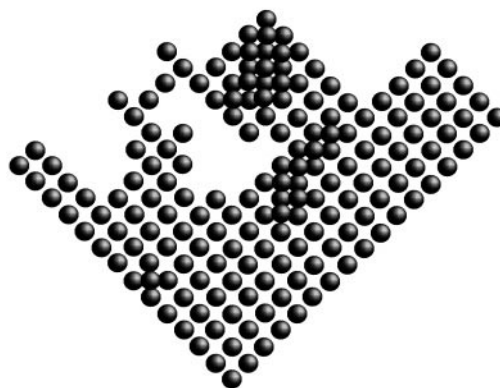
**Ion irradiation is a common technique of materials processing, as well as being relevant to the radiation damage incurred in nuclear reactors. Early models of the effects of ion irradiation typically assumed that particles undergo two-body elastic collisions<sup>1</sup>, like billiard balls colliding in three dimensions. Later descriptions invoked such phenomena as localization of kinetic energy, ther-**

**malization and localized melting<sup>2–4</sup>. In all these descriptions, the displacement of atoms is chaotic in that slight variations in the ion's trajectory produce completely different, unpredictable sets of atomic displacements<sup>5</sup>. Here we report molecular-dynamics simulations of high-energy self-bombardment of copper and nickel, in which we see collective displacements of atoms. The high pressures developed in collision cascades centred well below the surface can cause a coherent displacement of thousands of atoms, over tens of atomic planes, in a shear-induced slip motion towards the surface. The mechanism leads to a significant increase in damage production near the surface, characterized by well-ordered islands of adsorbed atoms. Our findings suggest an explanation for some features of radiation damage, as well as for differences between ion and neutron irradiation<sup>6</sup>.**

Much of what is currently known about ion and neutron irradiation phenomena in solids derives from binary-collision and molecular-dynamics (MD) computer simulations<sup>4</sup>. Binary-collision calculations have shown that an energetic ion impinging on a dense solid produces a sequence of chaotic atomic collisions in a small volume of a crystal, with the paths of recoiling atoms forming a fractal-like structure<sup>1</sup>. MD simulations, in contrast, show that a liquid-like zone can form in the centre of the collision cascade<sup>2</sup>, sometimes quenching into a strongly disordered amorphous phase<sup>7</sup>. In this picture atomic motion within cascades thus approaches a random walk comprised of ballistic and diffusive jumps.

The range of irradiation energies which can be examined by MD simulations has been limited by computer capacity. By using classical MD simulations with a linearly scalable decomposition of the simulation cell for massively parallel supercomputers we have now extended the simulations to a new energy range,  $E_{kin} \approx 50$ –200 keV (ref. 8). By systematically simulating collision cascades in Cu and Ni (produced by 50-keV energetic ions impinging on the surface or starting from the bulk), we show that in this energy range a coherent displacement process begins to become important. Simulations at higher energies are unlikely to reveal fundamentally different behaviour because above 200 keV the cascade structure becomes self-similar. In fact, less damage is created near the surface at high energies as the cross-sections for collisions decrease with energy. We chose Cu and Ni because they have almost the same atomic number, density and crystal structure (and are thus expected to have the same ballistic behaviour), but differ in elastic hardness and melting point.

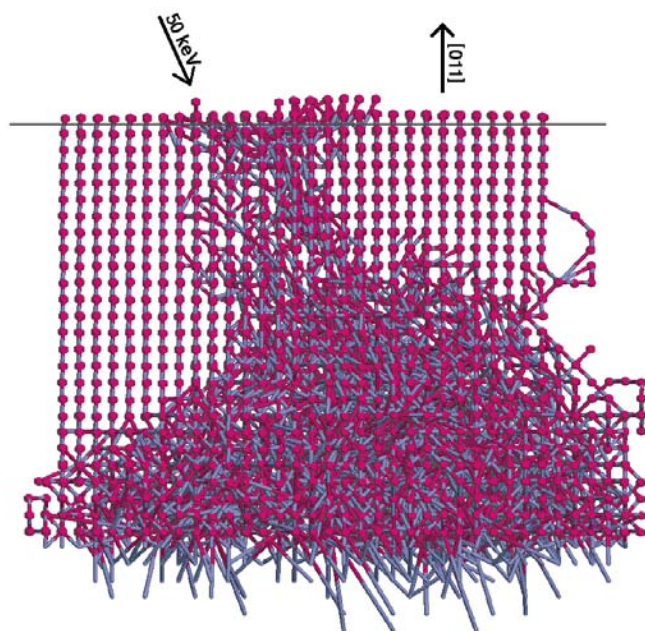
Our simulations show that the behaviour of surface events can be roughly divided into three categories. In the first, the ion penetrates deep into the sample, and the surface has little influence on damage



**Figure 1** Shape of adatom island ejected onto the surface by the shear mechanism in one 50-keV cascade in Cu. The straight edges of the adatom island are results of the coherent slip of atoms, as it is energetically favourable for the interstitial loop being forced outwards to be bounded by straight edge dislocations<sup>13</sup>.

production. In the second, hot liquid from the cascade flows or explodes onto the surface, producing adatoms, near-surface vacancies and nonlinear sputtering, as described previously<sup>9</sup>. These two events follow the traditional, chaotic view of cascade development. The third type of behaviour, however, differs dramatically from this traditional picture. It can occur when the cascade forms a liquid-like region that is sufficiently deep below the surface that the surface does not rupture directly. The extreme heat in the centre of the cascade produces a large compressive stress on the surrounding crystal, large enough to trigger the formation of an interstitial dislocation loop. The pressure can then cause planes of atoms to slide coherently along one of the  $\langle 110 \rangle$  slip directions of face-centred cubic (f.c.c.) metals towards the surface, forming an adatom island. The shearing probability is much larger close to the surface than in the bulk, owing to the elastic dipole interaction in the bulk between dislocation loops of opposite character; that is, between loops of vacancies or interstitials.

We recognized this 'shear mechanism' by tracking the motion of all atoms displaced from their initial lattice site over the cascade. In the shear events, adatoms—and the same number of atoms in several (001) planes below them—were all displaced exactly one nearest-neighbour distance along the same  $\langle 110 \rangle$  direction. We observed the shear mechanism once in seven Cu events and twice in seven Ni events.

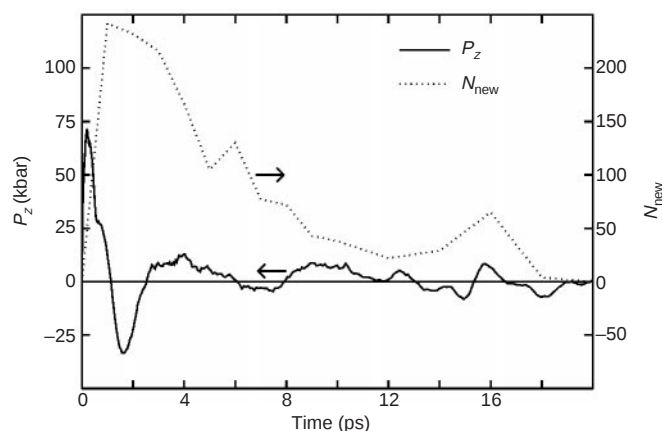


**Figure 2** Side view of atoms displaced by more than half a nearest-neighbour distance in a 50-keV Cu cascade event. Each line connects the initial and final position of one atom; the blue end of the line denotes the initial, and the red end of the line and the red sphere the final position of each atom. The black straight line shows the position of the surface, and the black arrows indicate the initial direction of the 50-keV incident ion and the  $[011]$  slip direction. The initial ion velocity was chosen in a non-channelling direction close to the  $[001]$  surface normal. Displaced atoms deeper than 45 Å have been omitted. Inside the sample, where the cascade forms a liquid region, the motion is essentially random, but near the surface most atoms have moved coherently towards the surface. The atom motion occurred in a  $\langle 110 \rangle$  direction and the side planes of the displaced atoms were  $\{111\}$  planes, as expected for the f.c.c. metal slip system. We also observed a few replacement collision sequences moving one-dimensionally along a single  $\langle 110 \rangle$  atom row<sup>14</sup>, but these have only a small effect on the overall cascade development<sup>15</sup>. The shear probability is expected to increase in materials with a lower shear modulus; indeed, preliminary results in the elastically softer materials Au and Pb show an even more pronounced effect, with atoms being sometimes displaced over two interatomic distances.

In the shear event in Cu, a large adatom island containing  $\sim 180$  adatoms formed mainly by the shear mechanism, although a few atoms in the same island were brought there by liquid flow. The event is illustrated in Figs 1–4. Figure 1 shows the final shape of the adatom island viewed from above. We note that two sides of the island form straight edges encompassing  $\sim 15$  atom rows. Such a well-defined form is very unlikely to have formed in the random, liquid flow-like events as a direct result of the collision cascade. Figure 2 illustrates the atom motion. It is evident that most of the adatoms are results of a uniform shear in the same  $\langle 110 \rangle$  direction. Analysis of the atom motion showed that at least 117 out of the 177 adatoms were part of the coherent  $\langle 110 \rangle$  motion. The other 60 adatoms were partly displaced randomly due to liquid flow along the ion path. The  $\langle 110 \rangle$  slip occurred over 20 atom layers in the edge region. In the two shear events in Ni  $\sim 30$  adatoms formed per event, and the slip was well-separated from liquid flow.

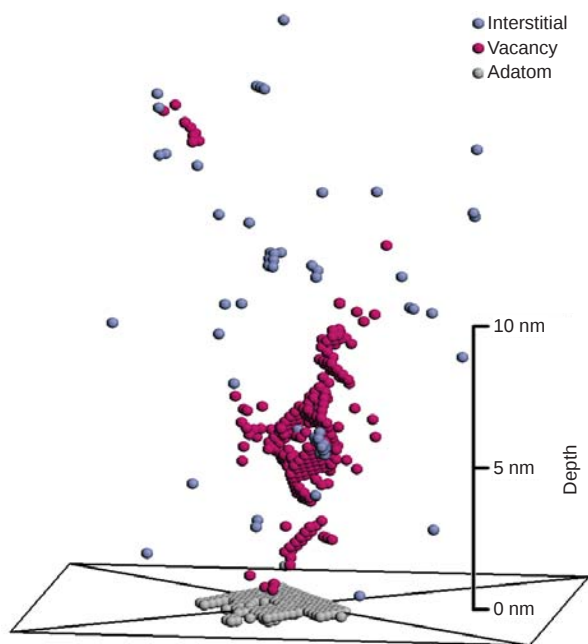
Figure 3 shows the development with time of the pressure and of the number of new nearest neighbours,  $N_{\text{new}}$ , of the atoms which are part of the adatom island. The time development of  $N_{\text{new}}$  shows that the adatom island forms mainly during the first 8 ps of the cascade, in agreement with animations of the cascade development. The shearing stress active during the formation of the adatom island is just at the threshold of producing a slip of atoms,  $\sim G/30$  (where  $G$  is the shear modulus)<sup>10</sup>. This explains why the slip is not observed in all events. For similar reasons, slip is not expected in cascades where the densest regions are deep inside the sample. In such deeply centred cascades the shear stress driving the dislocation loop away from the cascade becomes small before the attractive image force from the surface becomes significant.

We have shown that the contraction phase of the liquid zone formed by cascades inside the material results in efficient separation of interstitials and vacancies<sup>11</sup>, preventing their instant athermal recombination. When a surface is present, the separation becomes even more pronounced. The shear mechanism can produce adatom islands far from the liquid zone of the cascade. Because some atoms have thus escaped the central cascade region, a vacancy cluster is produced in the centre when the liquid recrystallizes. Because no damage is produced in the regions where atoms have slipped, the separation distance between the vacancies and adatoms is larger than in the bulk. The resulting defect distribution is illustrated in Fig. 4 for the Cu cascade discussed above. Most vacancies occur in one large cluster well below the surface, extending roughly from a



**Figure 3** Time development of  $P_z$  and  $N_{\text{new}}$  in the 50-keV Cu cascade forming the adatom island.  $P_z$  is the pressure in the direction of the surface normal of  $\sim 11,000$  atoms below the adatom island (including the adatoms), and  $N_{\text{new}}$  is the change in the number of nearest neighbours over a 1-ps interval. The latter was evaluated by finding all neighbouring atoms of each final adatom every 1 ps. If one of these neighbours was not a neighbour of the same adatom at the previous step, it was counted as a new neighbour.





**Figure 4** Side view of the final positions of all adatoms, interstitials and vacancies formed in the event forming the large adatom islands. Nine atoms were sputtered in this event. The black rectangle indicates the position of the surface. We note that the adatom and vacancy clusters are well-separated from each other, and that the vacancy cluster<sup>16</sup> already shows the rudiments of a stacking-fault tetrahedron (which is often observed in experiments<sup>17</sup>), even though no post-cascade annealing has taken place. In some other 50-keV events in Cu and Ni, we have observed the formation of perfect stacking fault tetrahedra.

depth of 5 to 10 nm. The crystal just below the surface has regenerated almost perfectly.

Because the energy required to produce adatoms (through either the liquid-flow or the shear mechanism) is much lower than the energy required to produce interstitials, damage production by cascades close to a surface is likely to be much larger than that produced by cascades in the bulk. Quantitative analysis of 7 cascades with a surface present and 5 cascades in the bulk in both Cu and Ni showed that the damage produced in surface events was on average  $\sim 4$  times the bulk average in Cu and 1.5 times the bulk value in Ni. The reason for the much smaller difference in Ni is that the surface effects are less likely to influence damage production. The higher melting point of Ni (compared to Cu) makes the liquid region smaller, reducing the probability of surface rupture and thus also reducing the probability of liquid flow. The higher shear modulus and smaller melt radius reduces the probability for slip of atom planes in Ni.

The strong enhancement of vacancy production in the vicinity of a surface explains why the damage observed (by electron microscopy) after ion implantation into metals is dominated by vacancy clusters<sup>12</sup>; it can also explain differences between damage produced by ion and neutron irradiation<sup>6</sup>. The effect of melting point and shear modulus on the surface enhancement of damage production is also a likely explanation of why more damage is observed in soft metals like Cu than in harder metals like Ni during irradiation close to the surface<sup>6</sup>. □

Received 30 June; accepted 4 December 1998.

1. Moreno Marin, J. C., Conrad, U. & Urbassek, H. M. Fractal structure of collision cascades. *Nucl. Instrum. Methods Phys. Res. B* **48**, 404–407 (1990).
2. Diaz de la Rubia, T., Averback, R. S., Benedek, R. & King, W. E. Role of thermal spikes in energetic collision cascades. *Phys. Rev. Lett.* **59**, 1930–1933 (1987).
3. Bacon, D. J. & Diaz de la Rubia, T. Molecular dynamics computer simulations of displacement cascades in metals. *J. Nucl. Mater.* **216**, 275–290 (1994).
4. Averback, R. S. & Diaz de la Rubia, T. in *Solid State Physics* Vol. 51 (eds Ehrenfest, H. & Spaepen, E.) 281–402 (Academic, New York, 1998).
5. Robinson, M. T. The statistics of sputtering. *Nucl. Instrum. Methods Phys. Res. B* **90**, 509–512 (1994).

6. English, C. A. & Jenkins, M. L. Insight into cascade processes arising from cascade collapse. *Mater. Sci. Forum.* **15–18**, 1003–1022 (1987).
7. Diaz de la Rubia, T. & Gilmer, G. H. Structural transformations and defect production in ion implanted silicon: a molecular dynamics simulation study. *Phys. Rev. Lett.* **74**, 2507–2510 (1995).
8. Nordlund, K., Wei, L., Zhong, Y. & Averback, R. S. Role of electron-phonon coupling on collision cascade development in Ni, Pd and Pt. *Phys. Rev. B* **57**, 13965–13968 (1998).
9. Ghaly, M. & Averback, R. S. Effect of viscous flow on ion damage near solid surfaces. *Phys. Rev. Lett.* **72**, 364–367 (1994).
10. Kittel, C. *Introduction to Solid State Physics* 3rd edn (Wiley, New York, 1968).
11. Nordlund, K. & Averback, R. S. Point defect movement and annealing in collision cascades. *Phys. Rev. B* **56**, 2421–2431 (1997).
12. Fukushima, H., Jenkins, M. L. & Kirk, M. A. On the determination of the nature of defect clusters produced by displacement cascades. Part II. Application of stereo imaging techniques to heavy-ion damage in copper. *Phil. Mag. A* **75**, 1583–1602 (1997).
13. Brown, L. M. & Woolhouse, G. R. The loss of coherency of precipitates and the generation of dislocations. *Phil. Mag.* **21**, 329–345 (1970).
14. Gibson, J. B., Goland, A. N., Milgram, M. & Vineyard, G. H. Dynamics of radiation damage. *Phys. Rev.* **120**, 1229–1253 (1960).
15. Nordlund, K. *et al.* Defect production in collision cascades in elemental semiconductors and FCC metals. *Phys. Rev. B* **57**, 7556–7570 (1998).
16. Seeger, A. in *Radiation Damage in Solids* Vol. 1, 101–127 (Int. Atomic Energy Agency, Vienna, 1962).
17. Kojima, S. *et al.* Confirmation of vacancy-type stacking fault tetrahedra in quenched, deformed and irradiated face-centred cubic metals. *Phil. Mag. A* **59**, 519–532 (1989).

**Acknowledgements.** This work was supported by the Academy of Finland, the US National Science Foundation and the US Department of Energy. Grants of computer time from the Center for Scientific Computing in Espoo, Finland, the National Energy Research Computer Center at Livermore, California, and the National Center for Supercomputing Applications in Champaign, Illinois, are gratefully acknowledged.

Correspondence and requests for materials should be addressed to K.N. (e-mail: kai.nordlund@helsinki.fi.).

## Two-photon polymerization initiators for three-dimensional optical data storage and microfabrication

Brian H. Cumpston<sup>\*</sup>, Sundaravel P. Ananthavel<sup>\*</sup>, Stephen Barlow<sup>†</sup>, Daniel L. Dyer<sup>†</sup>, Jeffrey E. Ehrlich<sup>‡</sup>, Lael L. Erskine<sup>†</sup>, Ahmed A. Heikal<sup>\*</sup>, Stephen M. Kuebler<sup>\*</sup>, I.-Y. Sandy Lee<sup>\*</sup>, Dianne McCord-Maughon<sup>†</sup>, Jinqui Qin<sup>†</sup>, Harald Röckel<sup>†</sup>, Mariacristina Rumi<sup>\*</sup>, Xiang-Li Wu<sup>\*</sup>, Seth R. Marder<sup>†‡</sup> & Joseph W. Perry<sup>\*‡</sup>

<sup>\*</sup> Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

<sup>†</sup> Beckman Institute, California Institute of Technology, Pasadena, California 91125, USA

<sup>‡</sup> Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

<sup>§</sup> Department of Chemistry, Wuhan University, Wuhan 430072, China

<sup>||</sup> Department of Chemistry, The University of Arizona, Tucson, Arizona 85721, USA

Two-photon excitation provides a means of activating chemical or physical processes with high spatial resolution in three dimensions and has made possible the development of three-dimensional fluorescence imaging<sup>1</sup>, optical data storage<sup>2,3</sup> and lithographic microfabrication<sup>4–6</sup>. These applications take advantage of the fact that the two-photon absorption probability depends quadratically on intensity, so under tight-focusing conditions, the absorption is confined at the focus to a volume of order  $\lambda^3$  (where  $\lambda$  is the laser wavelength). Any subsequent process, such as fluorescence or a photoinduced chemical reaction, is also localized in this small volume. Although three-dimensional data storage and microfabrication have been illustrated using two-photon-initiated polymerization of resins incorporating conventional ultraviolet-absorbing initiators, such photopolymer systems exhibit low photosensitivity as the initiators have small two-photon absorption cross-sections ( $\delta$ ). Consequently, this approach requires high laser power, and its widespread use remains impractical. Here we report on a class of  $\pi$ -conjugated compounds that exhibit large  $\delta$  (as high as  $1,250 \times 10^{-50} \text{ cm}^4 \text{ s per photon}$ ) and enhanced two-photon sensitivity relative to ultra-

violet initiators. Two-photon excitable resins based on these new initiators have been developed and used to demonstrate a scheme for three-dimensional data storage which permits fluorescent and refractive read-out, and the fabrication of three-dimensional micro-optical and micromechanical structures, including photonic-bandgap-type structures<sup>7</sup>.

We reported recently that molecules having the general structural motifs D- $\pi$ -D, D- $\pi$ -A- $\pi$ -D and A- $\pi$ -D- $\pi$ -A (where  $\pi$  is a  $\pi$ -conjugated bridge, D is a donor, and A is an acceptor) have large  $\delta$  relative to the parent unsubstituted conjugated molecules<sup>8</sup>. Our experimental and theoretical studies showed that  $\delta$  can be enhanced by increasing the conjugation length and the donor and acceptor strengths, and that this enhancement can be correlated with the degree of symmetric intramolecular charge transfer.

As the donor substituents make the D- $\pi$ -D molecules electron-rich, we recognized that after one- or two-photon photoexcitation these chromophores would be able to transfer an electron even to relatively weak acceptors, and that this process could be used to activate various chemical reactions, such as polymerization. For example, 4,4'-bis(*N,N*-di-*n*-butylamino)-*E*-stilbene (**1**; see Fig. 1 and Table 1) has a ground-state oxidation potential of -35 mV (relative to ferrocenium/ferrocene) and an energy of 3.1 eV for the first one-photon-allowed excited state,  $S_1$ , which together provide a large driving force for a photoinduced electron-transfer reaction. This driving force is also accessible via two-photon excitation into the higher-lying state  $S_2$  followed by rapid non-radiative decay to  $S_1$  (ref. 8). As  $\delta$  is large for this class of molecules, two-photon-induced electron transfer should be efficient.

Efficient electron transfer from photoexcited **1** to various electron acceptors was shown by (1) steady-state fluorescence quenching, (2) fluorescence lifetime shortening, and (3) the appearance of electronic absorption bands due to the radical cation of **1**, which has  $\lambda_{\text{max}} = 620$  nm. Relatively large bimolecular quenching rate constants,  $k_Q$ , were obtained for photoinduced intermolecular charge transfer reactions of **1** with  $C_{60}$  in toluene ( $k_Q = 8.8 \times 10^{10} \text{ M}^{-1}\text{s}^{-1}$ ) and with acrylates in acetonitrile ( $k_Q \approx 2 \times 10^{10} \text{ M}^{-1}\text{s}^{-1}$  and  $6 \times 10^{10} \text{ M}^{-1}\text{s}^{-1}$  for two commercial triacrylates, SR9008 and SR454, respectively, from Sartomer Co., Exton, Pennsylvania, USA). Additionally, we synthesized covalently linked chromophore-acceptor systems wherein one or more of the amine-bound alkyl groups of **1** was replaced with an electron acceptor (**2-5**, Fig. 1). For these systems, the fluorescence is quenched by fast intramolecular electron transfer. Whereas the fluorescence lifetime,  $\tau_f$ , of **1** in acetonitrile is 1.5 ns,  $\tau_f$  for **2**, **3** and **4** is 90 ps, < 10 ps and 35 ps, respectively. The steady-state fluorescence intensity of **5** in acetonitrile is reduced by a factor of 33 relative to **1**.

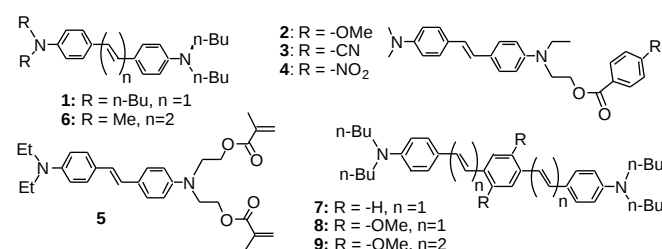
We reasoned that the radical ions formed by electron transfer from photoexcited **1** to an acrylate, or subsequent radical products, could initiate the polymerization of acrylates<sup>9</sup>. Indeed, polymerization of pure acrylate monomers (methylacrylate, methylmethacrylate and various tri- and penta-acrylates) was found to proceed via a radical mechanism following both one- and two-photon excitation of **1**. The growth rate ( $R_p$ ) of triacrylate polymer features photo-initiated by **1** was measured as a function of the incident laser intensity,  $I$ . For excitation into  $S_1$  at 355 nm,  $R_p$  depended on  $I^{1/2}$ , as

expected for one-photon-initiated polymerization<sup>9</sup>. In contrast, laser excitation into  $S_2$  at 600 nm led to a linear dependence of  $R_p$  on  $I$ , as one would predict for two-photon-initiated polymerization (TPIP) by extending the treatment described in ref. 9. Furthermore, for excitation across the two-photon absorption band the wavelength dependence of  $R_p$  was found to be similar to the two-photon excitation spectrum.

We compared  $R_p$  and the threshold energies ( $E_{\text{th}}$ ) for TPIP of SR9008 by **1** and **6** with those for the following conventional ultraviolet (UV) radical photoinitiators<sup>9</sup> under excitation at 600 nm with 5-ns pulses and a radial spot size of 225  $\mu\text{m}$ : benzil, benzophenone, bis(*N,N*-dimethylamino)benzil, isopropyl-thioxanthone, 1-[4-(methylthio)phenyl]-2-methyl-2-morpholinopropan-1-one (MP) and 4,4'-bis(*N,N*-dimethylamino)benzophenone (DABP). The most photosensitive UV initiators were MP and DABP, for which  $E_{\text{th}} = 1.0$  mJ. Compounds **1** and **6** had higher two-photon photosensitivity, with  $E_{\text{th}} = 0.3$  mJ, and  $R_p$  of **1** exceeding that of MP by a factor of ten. In general, the two-photon photosensitivity depends on both  $\delta$  and the overall chemical efficiency of initiation. We determined that the values of  $\delta$  at 600 nm are less than  $10 \times 10^{-50} \text{ cm}^4 \text{ s per photon}$  for the UV photoinitiators. These data indicate that the enhanced  $\delta$  of D- $\pi$ -D molecules contributes significantly to the improvement of the two-photon photosensitivity. We believe that any route to more photosensitive two-photon initiators should involve the design and synthesis of molecules possessing: (1) a chromophoric group with a large  $\delta$ , such as a D- $\pi$ -D structure; (2) a chemical functionality that has a high efficiency of initiation, such as those in UV initiators; and (3) a mechanism by which excitation of the chromophore leads to activation of the chemical functionality, such as an electron-transfer process.

High-density, three-dimensional (3D) optical data storage can be achieved by writing bits in a thick (>100  $\mu\text{m}$ ) storage medium using TPIP under tight-focusing conditions. Relative to a one-photon-based process, much higher information densities can be obtained by writing multiple layers of bits; this is because, first, the excitation light penetrates deeply into the material, being absorbed only at the focal region, and second, Rayleigh scattering is reduced for the longer wavelengths used for two-photon excitation. With this scheme, bit densities larger than  $10^{12} \text{ bits cm}^{-3}$  have been demonstrated<sup>2</sup>. However, this approach has not been commercially viable as the currently available initiators are weak two-photon absorbers, and therefore high-power excitation sources are required. The increased two-photon photosensitivity afforded by the D- $\pi$ -D structural motif provides a route to low-power, rapid writing in 3D optical data storage. We have further demonstrated<sup>8</sup> that for these systems, extending the conjugation length increases  $\delta$  ( $\delta = 1,250 \times 10^{-50} \text{ cm}^4 \text{ s per photon}$  for **9**). Chromophores **7-9** were also found to initiate polymerization of acrylates under two-photon excitation. In the case of **8**,  $E_{\text{th}}$  for 730-nm, 5-ns pulses was as low as 0.2 mJ.

We have used the D- $\pi$ -D molecules to demonstrate a means for 3D optical data storage based on fluorescent bits. As discussed above, **5** fluoresces weakly in comparison to **1**, largely because intramolecular electron transfer to the pendant acrylate groups



**Figure 1** Molecular structure of the compounds discussed in this work.

**Table 1** Optical and electrochemical data for D- $\pi$ -D chromophores

| Compound | $\lambda_{\text{max}}^{(1)}$ | $\lambda_{\text{max}}^{(2)}$ | $\delta$ | $E_{\text{D}^{\bullet+}/\text{D}}$ |
|----------|------------------------------|------------------------------|----------|------------------------------------|
| <b>1</b> | 374                          | 600                          | 210      | -35                                |
| <b>6</b> | 390                          | 645                          | 260      | -80*                               |
| <b>7</b> | 410                          | 730                          | 995      | +90                                |
| <b>8</b> | 425                          | 730                          | 900      | -10                                |
| <b>9</b> | 448                          | 775                          | 1,250    | -45                                |

$\lambda_{\text{max}}^{(1)}$  and  $\lambda_{\text{max}}^{(2)}$  are the one- and two-photon absorption maxima in nm, respectively,  $\delta$  is the peak two-photon absorption cross-section in  $10^{-50} \text{ cm}^4 \text{ s per photon}$ , and  $E_{\text{D}^{\bullet+}/\text{D}}$  is the oxidation potential in mV vs ferrocenium/ferrocene in THF / 0.1 M [ $\text{Bu}_4\text{N}$ ][PF<sub>6</sub>]. The two-photon absorption cross-sections were measured with nanosecond laser pulses using the apparatus described in ref. 8.

\* Two-electron oxidation observed.

provides a competitive de-excitation pathway. However, when **5** is polymerized, the C=C of the acrylate groups is consumed in the formation of the polymer backbone, greatly diminishing the electron-accepting ability and restoring the fluorescence. For example, the ratio of the fluorescence intensity for polymerized **5** versus unpolymerized **5** in acetonitrile is 7.3. We have used this approach to demonstrate optical data storage (Fig. 2). Fluorescent bits were recorded by TPIP in a semi-solid acrylate film containing **5**. The information was read out as a fluorescence image produced by exposing the entire film to low-intensity UV radiation. The written bits appear as bright spots on a dark background with a brightness contrast of  $\sim 3$ . As the crosslinking induced in these bits causes the refractive index to increase, the information can also be read out as a difference in refractive index between the polymerized and unpolymerized regions<sup>2</sup>, thus providing a second means of retrieving the stored data.

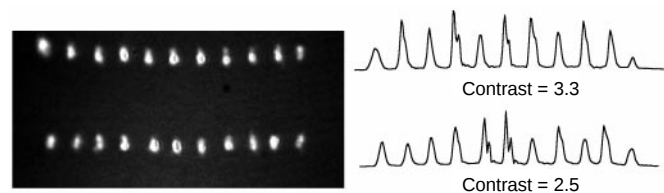
TPIP has also been utilized for 3D lithographic microfabrication (3DLM)<sup>4–6</sup>. In this process, a pattern is impressed into a photopolymer by translating the focus of an intense laser beam within the material. The localized photoexcitation crosslinks the photopolymer resin thereby reducing the solubility of the exposed material. The desired 3D structure is then obtained by dissolving away the unexposed material. This technology is promising, in that complex 3D microstructures, which would be either difficult or time-consuming to fabricate by any other technique, can be produced in a single step.

Having developed more efficient two-photon photoinitiators based on D- $\pi$ -D chromophores, we explored their use in 3DLM. Two-photon photopolymer resins were formulated, consisting of a polymer binder, a crosslinkable acrylate monomer, and D- $\pi$ -D chromophore **8** or **9** as the photoinitiator. Films of these resins were exposed to highly focused laser pulses ( $\sim 0.35$ - $\mu\text{m}$  radial spot size, 150-fs duration, at a repetition frequency of 76 MHz) at a wavelength near the two-photon absorption peak of the initiator. The threshold power for writing ( $P_{\text{th}}$ ) was 200  $\mu\text{W}$  for **8** at 730 nm and 300  $\mu\text{W}$  for **9** at 800 nm. In both cases, damage was observed at powers greater than  $\sim 10$  mW. The dynamic power range for polymerization is then 50 and 33 for **8** and **9**, respectively. In the absence of initiator, no writing occurred at powers below the damage threshold. For comparison, films in which the photoinitiator was either MP, benzil, or DABP were found to have dynamic ranges varying from  $\sim 1.4$  to 4 at 730 nm. Thus, use of **8** or **9** has resulted in photopolymers with an order-of-magnitude improvement in two-photon photosensitivity, relative to the conventional initiators examined.

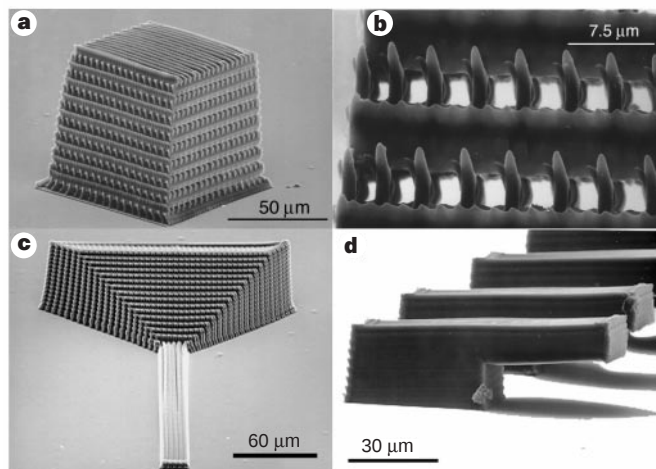
To illustrate the utility of our two-photon photopolymer systems, we used the 3DLM approach to produce a variety of potentially interesting 3D structures. Three-dimensional periodic structures are needed for photonic bandgap (PBG) materials, which have unique optical properties<sup>7</sup>. These structures are difficult to fabricate

on micrometre or submicrometre length scales as needed for applications in the infrared and visible spectral regions<sup>10</sup>. Figure 3a and b shows views of a 'stack-of-logs' structure which, given a sufficiently high refractive index, should exhibit PBG properties in the infrared spectral region. High-refractive-index structures, based for example on refractory ceramics, could be fabricated by using the polymer structure as a preform<sup>10</sup>. Other structures of arbitrary pattern and periodicity down to  $\sim 1$   $\mu\text{m}$  length scales could be readily obtained with this photopolymer system. Two-photon 3DLM could also find application in the production of tapered optical waveguides, such as those shown in Fig. 3c. For the example shown, the waveguide cross-section varies along the length from a 100  $\mu\text{m} \times 100$   $\mu\text{m}$  square aperture to a 2  $\mu\text{m} \times 10$   $\mu\text{m}$  rectangular aperture. Tapered optical waveguides have the potential to reduce optical loss in the coupling of waveguide components with disparate cross-sections<sup>11</sup>. Microelectromechanical systems (MEMS) are often produced using two-dimensional lithography, and 3D structures are built up by an iteration of processing steps<sup>12</sup>. As an example of MEMS structures easily fabricated with a single development step, we produced the array of polymeric cantilevers shown in Fig. 3d.

This work demonstrates that symmetric, donor-substituted conjugated chromophores designed to have large  $\delta$  and strong electron-donating properties are effective as two-photon-excitabile charge-transfer agents and are able to initiate polymerization of acrylate monomers. The results described here open a path towards the design of chromophores with even higher efficiency for the activation of radical polymerization. These systems would offer lower polymerization thresholds, higher polymerization rates, and wider dynamic write ranges, all of which are critical in 3D data storage and microfabrication, where fast data recording, rapid fabrication, and low-power writing are essential. Moreover, these principles could be applied equally well in the design of properly functionalized two-photon absorbers, which are engineered to activate other chemical processes, thereby making efficient 3D optical data storage and lithographic microfabrication possible with other materials systems. □



**Figure 2** Fluorescent bits recorded by two-photon-initiated polymerization. Left, contrast-enhanced fluorescence image of two rows of recorded bits. Right, intensity profile of the fluorescence from the two arrays of bits displayed on the left. The storage medium was 1% **5**, 49% SR9008 (inhibitor removed), and 50% poly(styrene-co-acrylonitrile) (75:25) as a polymer binder. Bits were written using focused 600-nm, 4-ps pulses at a repetition rate of 3.8 MHz with a spot size of 1.1  $\mu\text{m}$ , average power of 30 mW, and exposure times ranging from 0.01 to 1 s. The fluorescence image was recorded by irradiating the sample with low-intensity UV radiation.



**Figure 3** Three-dimensional microstructures produced by two-photon-initiated polymerization. **a**, Photonic bandgap structure. **b**, Magnified top-view of structure in **a**. **c**, Tapered waveguide structure. **d**, Array of cantilevers. The resins consisted of 0.1% photoinitiator (**8** or **9**), 70% reactive trifunctional acrylate monomer (typically Sartomer SR9008 and SR368), and 29.9% poly(styrene-co-acrylonitrile) (75:25) as a polymer binder, with a radical inhibitor level of  $\sim 30$  p.p.m. Films with a thickness of 120–150  $\mu\text{m}$  were formed by solvent evaporation from a dioxane solution. The films were exposed to Ti:sapphire laser pulses (150 fs, 76 MHz,  $\sim 0.35$   $\mu\text{m}$  radial spot size) at a wavelength near the two-photon absorption peak of the initiator. The substrate was translated at 50  $\mu\text{m s}^{-1}$  in a programmed 3D pattern during the exposure. The films were developed with an *N,N*-dimethylformamide wash.



Received 2 November 1998; accepted 11 January 1999.

1. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
2. Strickler, J. H. & Webb, W. W. Three-dimensional optical data storage in refractive media by two-photon point excitation. *Opt. Lett.* **16**, 1780–1782 (1991).
3. Parthenopoulos, D. A. & Rentzepis, P. M. 3-Dimensional optical storage memory. *Science* **245**, 843–845 (1989).
4. Strickler, J. H. & Webb, W. W. Two-photon excitation in laser scanning fluorescence microscopy. *SPIE Proc.* **1398**, 107–118 (1990).
5. Wu, E. S., Strickler, J. H., Harrell, W. R. & Webb, W. W. Two-photon lithography for microelectronic application. *SPIE Proc.* **1674**, 776–782 (1992).
6. Maruo, S., Nakamura, O. & Kawata, S. Three dimensional microfabrication with two-photon-absorbed photopolymerization. *Opt. Lett.* **22**, 132–134 (1997).
7. Yablonovitch, E. Photonic crystals. *J. Mod. Opt.* **41**, 173–194 (1994).
8. Albota, M. *et al.* Design of organic molecules with large two-photon absorption cross sections. *Science* **281**, 1653–1656 (1998).
9. Odian, G. *Principles of Polymerization* 2nd edn (Wiley, New York, 1981).
10. Feiertag, G. *et al.* Fabrication of photonic crystals by deep x-ray lithography. *Appl. Phys. Lett.* **71**, 1441–1443 (1997).
11. Wu, L., Li, F., Tang, S., Bihari, B. & Chen, R. T. Compression-molded three-dimensional tapered polymeric waveguides for low-loss optoelectronic packaging. *IEEE Photon. Technol. Lett.* **9**, 1601–1603 (1997).
12. Bustillo, J. M., Howe, R. T. & Muller, R. S. Surface micromachining for microelectromechanical systems. *Proc. IEEE* **86**, 1552–1574 (1998).

**Acknowledgements.** We thank S. Fraser and the Biological Imaging Center, California Institute of Technology, for use of a two-photon scanning laser microscope system, and S. Thayumanavan for synthetic precursors used in this work. H.R. thanks the Alexander von Humboldt Stiftung for a postdoctoral fellowship through the Theodore Lynen Program. This work was supported by the Ballistic Missile Defense Initiative Organization (at the Jet Propulsion Laboratory) and by the Office of Naval Research (through CAMP at the University of Arizona), the NSF and the Air Force Office of Scientific Research (at the California Institute of Technology).

Correspondence and requests for materials should be addressed to J.W.P. (e-mail: jwperry@u.arizona.edu).

## Photocontrolled nanophase segregation in a liquid-crystal solvent

Yves Lansac<sup>\*†</sup>, Matthew A. Glaser<sup>\*</sup>, Noel A. Clark<sup>\*</sup> & Oleg D. Lavrentovich<sup>‡</sup>

<sup>\*</sup> Condensed Matter Laboratory, Department of Physics, and Ferroelectric Liquid Crystal Materials Research Center, University of Colorado, Boulder, Colorado 80309, USA

<sup>‡</sup> Liquid Crystal Institute and Chemical Physics Program, Kent State University, Kent, Ohio 44242, USA

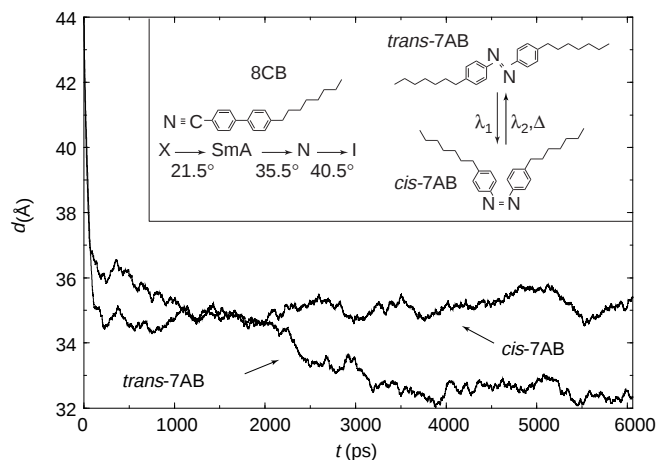
Materials whose structure or electrical or optical properties can be controlled with light are said to be photoactive. Liquid crystals are of interest as photoactive media, because their fluidity maintains the possibility of molecular motion in response to photon absorption, while their orientational and/or positional ordering offers the possibility of cooperative behaviour that can amplify relatively weak photochemical effects. Moreover, liquid crystals impose their own ordering on solutes. For example, smectic A liquid crystals, comprised of one-dimensional stacks of fluid layers with the molecular axes aligned normal to the layers<sup>1</sup>, produce a modulation in solute concentration with a period equal to the layer spacing<sup>2,3</sup>. Here we present computer simulations which show that the positional ordering of a photoactive solute (an azobenzene derivative, denoted 7AB) in a smectic host (denoted 8CB) depends sensitively on its photochemical state. The photoactive molecules are driven from within the smectic layers to locations between the layers by *trans*-to-*cis* photoisomerization. This would explain the recent observation<sup>4–7</sup> of a reversible increase in the smectic A layer spacing of a solution of 7AB in 8CB accompanying the photoisomerization process. The effect might be exploited for low-power, high-resolution optical data storage, and more generally for the manipulation of organic materials at the nanometre scale.

<sup>†</sup> Present address: Dipartimento di Chimica Fisica e Inorganica, Università di Bologna, Viale Risorgimento 4, 41036 Bologna, Italy.

To investigate the microscopic origins of the observed photo-mechanical response of mixtures of 7AB (*p,p'*-diheptylazobenzene) in 8CB (4-octyl-4'-cyanobiphenyl), and to probe directly the positional, orientational, and conformational distributions of *trans*- and *cis*-7AB and 8CB in smectic A solutions (Fig. 1), we have performed three separate molecular-dynamics simulations: one of pure 8CB, one of a *trans*-7AB/8CB mixture, and one of a *cis*-7AB/8CB mixture. The simulations were performed at constant pressure and temperature<sup>8</sup>, under thermodynamic conditions that correspond closely to those of the experiment (pressure, 1 atm; temperature, 17 °C). Periodic boundary conditions were employed, using a unit cell of variable size and shape to enable both the layer spacing and the in-layer areal density of molecules to adjust to their equilibrium values. All simulated systems contained 100 8CB molecules, with the 7AB/8CB mixtures additionally containing 12 7AB molecules (10.7 mol% 7AB), a composition close to that studied experimentally (9.6 mol% 7AB). Relatively long runs (in excess of 6 ns) were performed to verify that smectic ordering was stable, and to ensure that an equilibrium distribution of 7AB molecules was obtained.

We utilized a hybrid molecular model, in which *sp*<sup>3</sup>-hybridized CH<sub>n</sub> groups were treated as spherically symmetric united atoms. The interaction potential, derived from both *ab initio* and empirical input<sup>9</sup>, contains valence (bond stretch, bond angle bend, and dihedral torsion) interactions as well as non-bonded (van der Waals and Coulomb) site–site interactions. Long-range Coulomb interactions were evaluated to high accuracy using the particle-mesh Ewald method<sup>10</sup>, and a standard long-range correction for van der Waals interactions was employed<sup>11</sup>. The computational cost of the simulations was significantly reduced by employing a stable, highly optimized multiple-timestep molecular-dynamics integrator<sup>12,13</sup>.

In light of the partial bilayer structure of smectic A 8CB (with a layer spacing of *d* = 31.432 Å, ~1.4 times the fully extended molecular length of 22.1 Å), we used a bilayer initial condition for 8CB in all three simulations, with a layer spacing of 44 Å and with both 7AB and 8CB molecules orientated with their long axes perpendicular to the plane of the bilayer. In the 7AB/8CB mixtures, the 7AB molecules were initially placed within the smectic layers. To preclude the possibility of a non-uniform distribution of 7AB molecules over smectic layers, the monoclinic unit cell is chosen

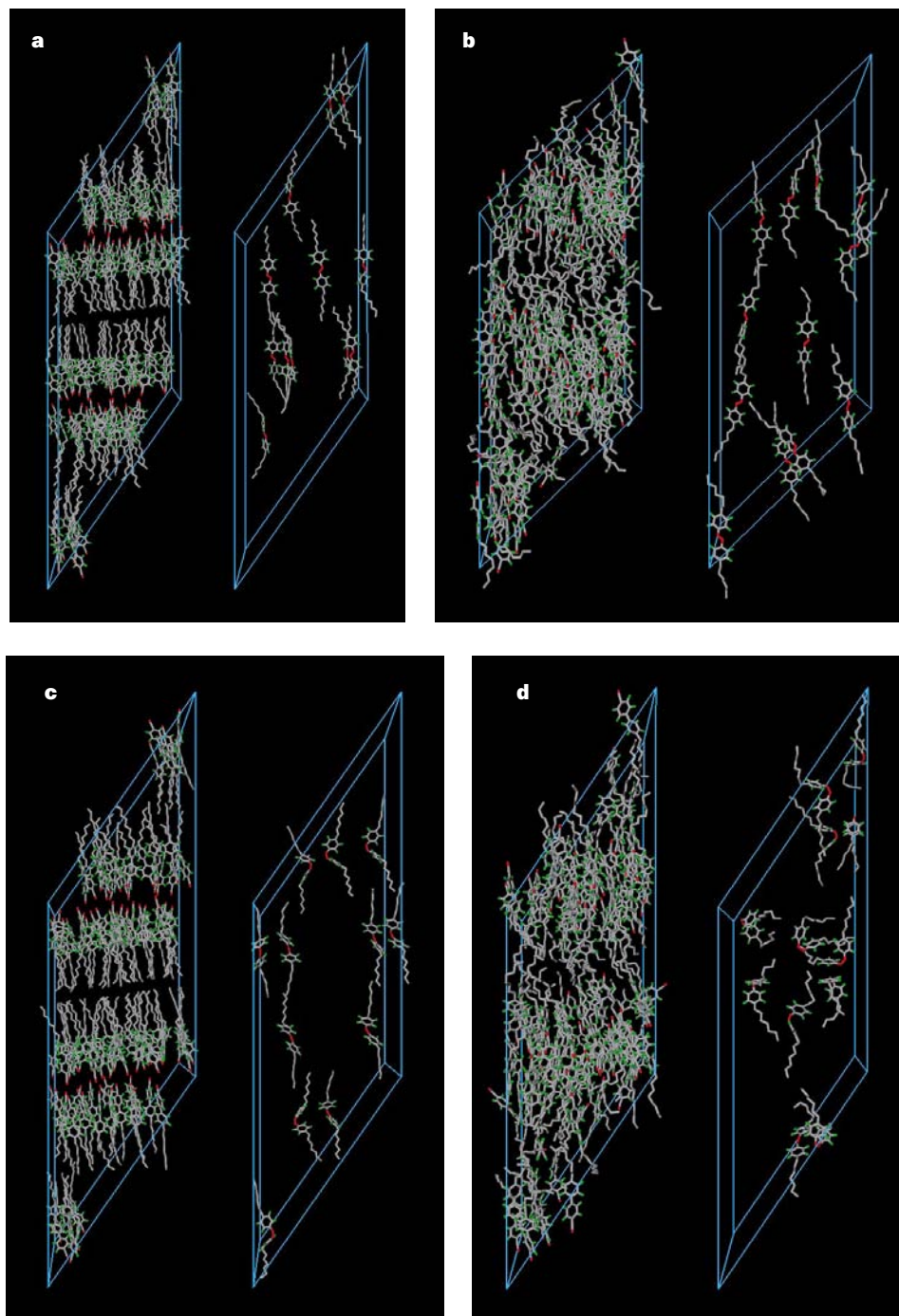


**Figure 1** Time evolution of smectic layer spacing *d* for the *trans*-7AB and *cis*-7AB simulations. Inset, chemical structures of 8CB (left) and *trans*- and *cis*-7AB (right). The phase transitions of pure 8CB are also shown. In the absence of illumination, *trans*-7AB is ~50 kJ mol<sup>-1</sup> more stable than the *cis* isomer, but can be converted to the *cis* form by electronic excitation under UV irradiation. The reverse reaction proceeds thermally or on illumination with longer-wavelength light.

so that the system effectively consists of a single continuous bilayer, taking into consideration the periodic boundary conditions. The initial configurations for the *trans* and *cis* simulations are shown in Fig. 2a and c, respectively.

The microscopic origin of the observed photomechanical effect in 7AB/8CB mixtures is apparent from Fig. 2b and d, which show the final configurations of the *trans*- and *cis*-7AB simulations, respectively, and from the average mass density profiles for 8CB and 7AB along the normal to the smectic layers (Fig. 3). A pronounced difference in the distribution of 7AB molecules in the two simulations is evident: *trans*-7AB molecules, which are essentially linear,

remain largely integrated within the 8CB layers, whereas bent-core *cis*-7AB molecules have been expelled from (and are distributed between) the 8CB layers. The change in layer structure produced by *trans*  $\rightarrow$  *cis* photoisomerization of 7AB is indicated schematically in Fig. 3. This difference in molecular organization is reflected in the evolution of smectic layer spacing  $d$  with time for the two simulations (Fig. 1). The time evolution of  $d$  for both systems is similar up to  $\sim 2$  ns, after which the *trans*-7AB layer spacing decreases rapidly to equilibrate at a value significantly lower than that for the *cis*-7AB simulation. The layer spacing and other thermodynamic quantities reach a steady state after  $\sim 4$  ns, indicating that thermodynamic



**Figure 2** Instantaneous configurations from large-scale atomistic simulations of 7AB-8CB mixtures. The initial ( $t = 0$  ns) and final ( $t = 6.06$  ns) configurations of the *trans*-7AB simulation are shown in **a** and **b**, respectively; the initial and final configurations of the *cis*-7AB simulation are shown in **c** and **d**. For

clarity, 8CB and 7AB molecules are displayed separately, respectively left and right in each figure part. The spatial segregation of *cis*-7AB between the smectic layers is seen in **d**, while *trans*-7AB is largely incorporated into the smectic layers (**b**).

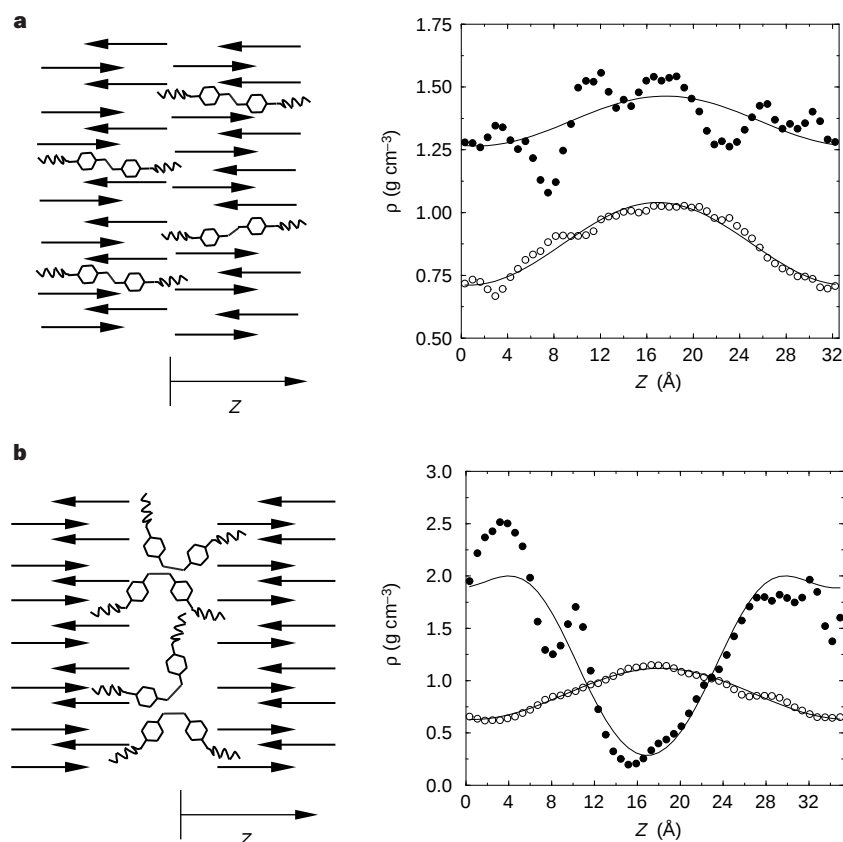
equilibrium is achieved within this time, so thermodynamic averages were calculated over the final 2 ns of each run.

The average layer spacing of the *cis*-7AB system is  $d_{\text{cis}} = 35.3 \text{ \AA}$ , while that of the *trans*-7AB system is  $d_{\text{trans}} = 32.6 \text{ \AA}$ , giving a layer spacing increase upon *trans*  $\rightarrow$  *cis* photoisomerization of  $\Delta d_{\text{sim}} = d_{\text{cis}} - d_{\text{trans}} = 2.7 \text{ \AA}$ . This is a much larger effect than that observed in experiments<sup>6,7</sup>, for which  $d_{\text{vis}} = 30.119 \text{ \AA}$  and  $d_{\text{UV}} = 30.274 \text{ \AA}$  (where  $d_{\text{vis}}$  and  $d_{\text{UV}}$  are the layer spacings measured in the absence and presence of UV illumination, respectively) giving  $\Delta d_{\text{exp}} = d_{\text{UV}} - d_{\text{vis}} = 0.155 \text{ \AA}$ , a factor of 17 smaller than  $\Delta d_{\text{sim}}$ . The discrepancy between simulation and experiment evidently reflects the fact that only a small fraction ( $f \approx 6\%$ ) of the *trans* isomers of 7AB are converted to the *cis* form under illumination by 366-nm UV radiation, based on  $^2\text{H}$  NMR studies of a 7AB/8CB mixture containing selectively deuterated 7AB (B. Zalar and D. Finotello, unpublished results), whereas our simulation effectively assumes  $f = 100\%$ . In fact, extrapolating  $\Delta d_{\text{exp}}$  to  $f = 100\%$  (assuming a linear dependence of  $\Delta d$  on  $f$ ) gives  $\Delta d_{\text{exp}} = 2.6 \text{ \AA}$ , in remarkably good agreement with simulation. Apparently, smectic ordering inhibits *trans*  $\rightarrow$  *cis* isomerization of 7AB, as  $f \approx 70\%$  is measured in isotropic solution<sup>5</sup>. We note that  $d_{\text{trans}}$  is more than  $2 \text{ \AA}$  larger than  $d_{\text{vis}}$ , a discrepancy probably due to the limitations of the intermolecular potential used in the simulations. For example, induction interactions (interactions between permanent and induced charge distributions) are not included in our model.

We computed orientational distributions of unit vectors attached to selected portions of the 7AB molecule, as shown in Fig. 4. The distributions for *trans*-7AB are as expected for linear molecules

incorporated into the 8CB layers, with alkyl tails preferentially orientated parallel to the layer normal, the N=N bond inclined by  $\sim 60^\circ$  with respect to the layer normal, and the normal to the C=N=N-C plane parallel to the smectic layers. The distributions for *cis*-7AB are also quite structured, with the normal to the C=N=N-C plane preferentially orientated parallel to the layers, and with N=N bonds orientated parallel to the layer normal. Moreover, the alkyl tails of *cis*-7AB are predominantly orientated along the layer normal, implying that a sizeable fraction of 7AB molecules have their alkyl tails interdigitated with the host 8CB matrix (Fig. 3). Thus, although *cis*-7AB molecules are expelled from the 8CB layers, their orientational distribution is still strongly conditioned by the 8CB matrix. The surprising observation that the N=N bond of *cis*-7AB is preferentially orientated along the layer normal is consistent with  $^2\text{H}$  NMR studies of 7AB/8CB mixtures containing selectively deuterated 7AB (B. Zalar and D. Finotello, unpublished results).

Our observations on 7AB/8CB systems have implications with regard to recent measurements on ferroelectric liquid crystals, tilted smectic (smectic C) phases containing chiral molecules, in which one of the chiral constituents is a photoactive azobenzene derivative. These experiments show a striking decrease in ferroelectric polarization density on exposure to UV radiation, and demonstrate that switching driven by a photochemical process is possible in such materials<sup>14–19</sup>. Our simulations suggest that the observed photo-modulation of ferroelectric polarization involves photocontrolled nanophase segregation of the photoactive component. Chiral photoactive molecules in the *trans* configuration are incorporated into the smectic C layers and contribute to the overall ferroelectric



**Figure 3** Mass density profiles along the layer normal for 8CB (open circles) and 7AB (filled circles), averaged over the final 2 ns of the *trans*-7AB (a) and *cis*-7AB (b) simulations. The mass densities for 7AB have been multiplied by a factor of 10 for clarity. The solid curves are fits to a low-order Fourier expansion. Whereas *trans*-7AB is preferentially incorporated into the 8CB layers (there is a weak maximum in the mass density profile for *trans*-7AB that coincides with the centre

of the 8CB partial bilayer), the density profile for *cis*-7AB exhibits a deep minimum near the centre of the 8CB layer, indicating that *cis*-7AB is quite effectively excluded from the 8CB layers. In relative terms, the density modulation for *cis*-7AB is much more pronounced than that for 8CB itself. The distribution of 7AB molecules within the smectic A 8CB matrix in the two cases is indicated schematically at left, with 8CB molecules represented by arrows.



polarization of the material by virtue of the polar orientational order imparted to them by the smectic C host<sup>20</sup>. On *trans* → *cis* photoisomerization, however, the *cis* isomers are expelled from the layers, and their contribution to the ferroelectric polarization decreases to near zero, as they are much less strongly orientated by the smectic C medium than *trans* isomers.

In general terms, this work demonstrates nanoscale photocontrol of the positional distribution of a photoactive solute in a smectic medium. The same effect should enable photocontrol of solute diffusion (photochemical pumping) and nanoscale photo-induced concentration of reactive species in smectic materials. Photocontrolled nanophase segregation should have a broad range of potential applications in nanotechnology and in the creation of novel photofunctional materials. Moreover, similar effects may play a role in the photoelectric response of bilayer lipid membranes containing azobenzene chromophores<sup>21</sup>.

We are, at present, unable to identify the specific energetic and/or entropic effects responsible for interlamellar nanophase segregation of *cis*-7AB in 7AB/8CB mixtures. A detailed calculation of distinct contributions to the free energy of a solute molecule (such as 7AB) in a smectic host as a function of position relative to the smectic layers is technically feasible, but would require at least an order of magnitude more computational effort than the present study. Understanding the microscopic mechanisms responsible for the formation of smectic phases is itself an unsolved problem, whose solution faces similar technical challenges, and which similarly falls

into a quite broad class of problems related to the self-assembly of soft condensed matter. Studies such as the present one represent first steps towards addressing these problems in the context of liquid-crystal self-assembly. □

Received 17 July; accepted 14 December 1998.

- De Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals* (Clarendon, Oxford, 1993).
- Rieker, T. P. Organic lyotropic lamellar liquid crystals. *Liq. Cryst.* **19**, 497–500 (1995).
- Guymon, C. A. *et al.* Effects of monomer structure on their organization and polymerization in a smectic liquid crystal. *Science* **275**, 57–59 (1997).
- Folks, W. R. *et al.* Low-power laser induced instabilities in smectic A liquid crystals. *Mol. Cryst. Liq. Cryst.* **261**, 259–270 (1995).
- Folks, W. R., Reznikov, Yu. A., Yarmolenko, S. N. & Lavrentovich, O. D. Light-induced periodic lattice of defects in smectic A and C liquid crystals: structural and dynamical aspects. *Mol. Cryst. Liq. Cryst.* **292**, 183–197 (1997).
- Krentsel, T. A., Lavrentovich, O. D. & Kumar, S. In-situ x-ray measurements of light-controlled layer spacing in a smectic liquid crystal. *Mol. Cryst. Liq. Cryst.* **304**, 463–469 (1997).
- Folks, W. R. *et al.* Photocontrol of smectic spacing. *Mol. Cryst. Liq. Cryst.* **320**, 77–88 (1998).
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
- Glaser, M. A., Clark, N. A., Garcia, E. & Walba, D. M. Quantum chemistry based force fields for soft matter. *Spectrochim. Acta A* **53**, 1325–1340 (1997).
- Essman, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
- Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Oxford Univ. Press, 1987).
- Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
- Sexton, J. C. & Weingarten, D. H. Hamiltonian evolution for the hybrid Monte Carlo algorithm. *Nucl. Phys. B* **380**, 665–677 (1992).
- Ikeda, T., Sasaki, T. & Ichimura, K. Photochemical switching of polarization in ferroelectric liquid-crystal films. *Nature* **361**, 428–430 (1993).
- Sasaki, T., Ikeda, T. & Ichimura, K. Photochemical control of properties of ferroelectric liquid crystals: photochemical flip of polarization. *J. Am. Chem. Soc.* **116**, 625–628 (1994).
- Sasaki, T. & Ikeda, T. Photochemical control of properties of ferroelectric liquid crystals. 1. Effect of structure of host ferroelectric liquid crystals on the photochemical switching of polarization. *J. Phys. Chem.* **99**, 13002–13007 (1995).
- Sasaki, T. & Ikeda, T. Photochemical control of properties of ferroelectric liquid crystals. 2. Effect of the structure of guest photoresponsive molecules on the photochemical switching of polarization. *J. Phys. Chem.* **99**, 13008–13012 (1995).
- Sasaki, T. & Ikeda, T. Photochemical control of properties of ferroelectric liquid crystals. 3. Photochemically induced reversible change in spontaneous polarization and electrooptic property. *J. Phys. Chem.* **99**, 13013–13018 (1995).
- Coles, H. J., Walton, H. G., Guillon, D. & Poetti, G. Photomechanically induced phase transitions in ferroelectric liquid crystals. *Liq. Cryst.* **15**, 551–558 (1993).
- Glaser, M. A. *et al.* Computer-aided design of ferroelectric liquid crystals. *Mol. Phys. Rep.* **10**, 26–47 (1995).
- Yamaguchi, H. & Nakanishi, H. Characterization of the preparation process and the photochemical control of electrical properties of bilayer lipid membranes containing azobenzene chromophores. *Biochim. Biophys. Acta* **1148**, 179–184 (1993).

**Acknowledgements.** We thank T. Darden for use of his particle-mesh Ewald code, D. Coleman for help in producing the molecular graphics, and D. Finotello and B. Zalar for discussions. This work was supported by the NSF, the US Air Force Office of Sponsored Research, and the Kent State Liquid Crystal Institute (a grant of computer time).

Correspondence and requests for materials should be addressed to M.A.G. (e-mail: matthew.glaser@colorado.edu).

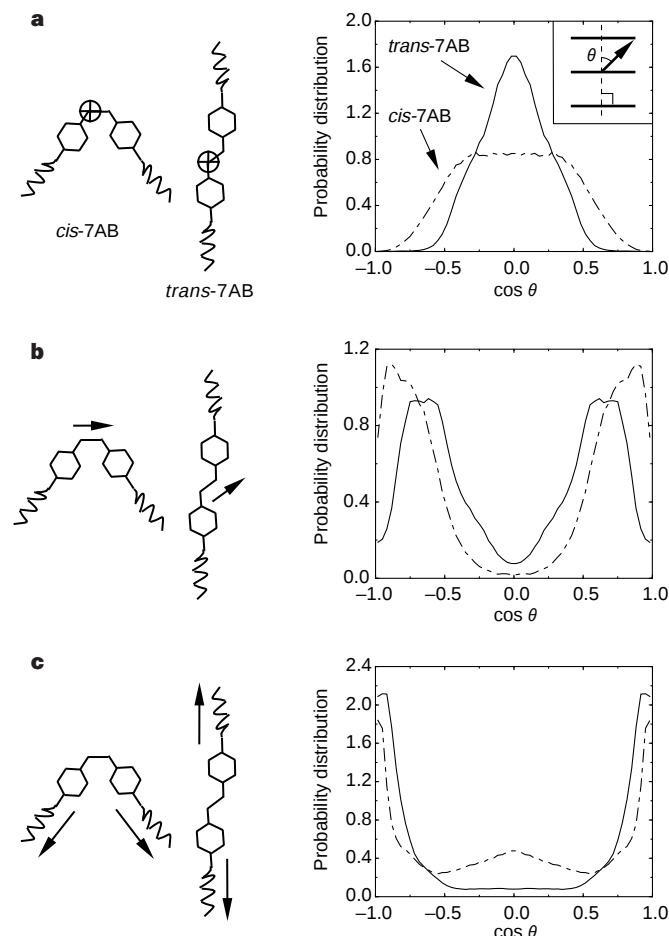
## The role of mat-forming diatoms in the formation of Mediterranean sapropels

Alan E. S. Kemp\*, Richard B. Pearce\*, Itaru Koizumi†, Jennifer Pike\*‡ & S. Jae Rance\*

\* School of Ocean and Earth Science, University of Southampton, Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK

† Division of Earth and Planetary Sciences, Graduate School of Science, Hokkaido University, Sapporo, 060, Japan

The origins of sapropels (sedimentary layers rich in organic carbon) are unclear, yet they may be a key to understanding the influence of climate on ocean eutrophication, the mechanisms of sustaining biological production in stratified waters and the genesis of petroleum source rocks<sup>1–3</sup>. Recent microfossil studies of foraminifera<sup>1</sup> and calcareous nannofossils<sup>2</sup> have focused attention on a deep chlorophyll maximum as a locus for the high



**Figure 4** Orientational distributions for selected portions of the 7AB molecule. Results are shown in terms of the cosine of the angle between a specific intramolecular vector and the layer normal. The distributions for a vector perpendicular to the C–N=N–C plane, for a vector parallel to the N=N bond, and for the end-to-end vectors of the alkyl chains are shown in **a**, **b** and **c**, respectively. Distributions for both *trans*-7AB (solid line) and *cis*-7AB (dot-dashed line) are shown.

‡ Present address: Department of Earth Sciences, University of Wales, Cardiff, PO Box 914, Cardiff CF1 3YE, UK.

production inferred<sup>3</sup> for sapropel formation, but have not identified the agent responsible. Here we report the results of a high-resolution, electron-microscope-based study of a late Quaternary laminated sapropel in which the annual flux cycle has been preserved. We find that much of the production was by diatoms, both mat-forming and other colonial forms, adapted to exploit a deep nutrient supply trapped below surface waters in a stratified water column. Reconstructed organic-carbon and opal fluxes to the sediments are comparable to those at high-productivity sites in today's oceans, and calculations based on diatom Si/C ratios suggest that the high organic-carbon content of sapropels may be entirely accounted for by sedimenting diatoms. We propose that this style of production may have been common in ancient Palaeogene and Cretaceous seas, environments for which conventional appeals to upwelling-driven production to account for the occurrence of diatomites, and some organic-carbon-rich sediments, have never seemed wholly appropriate.

The Mediterranean Sea has provided a focus for the contentious debate over the relative importance of stratification-driven anoxia versus productivity in the origins of organic-carbon-rich sediments. An ingredient that has been largely missing from this debate are inferences from the ecology of diatoms, the world's dominant marine primary producer. This is because the waters of the Mediterranean are highly undersaturated in silica, and the opaline skeletons of siliceous plankton are preserved only rarely in anoxic brine-basins whose waters are buffered and contain higher levels of dissolved silica<sup>4</sup>. Leg 160 of the Ocean Drilling Program (ODP) collected a suite of cores from the eastern Mediterranean that included a critical site (971) in the deepest part of the moat of the Napoli mud volcano on the Mediterranean ridge (Fig. 1). This natural sediment trap has preserved a late Quaternary laminated diatomaceous sapropel (S5) which records the annual flux cycle. We combine a detailed electron-microscope investigation of this sapropel (and diatom micropalaeontology of other sapropels cored during ODP Leg 160) with recent developments in the understanding of the ecology of mat-forming diatoms<sup>5</sup>, and of their significant role in generating massive flux<sup>6,7</sup>, to present a model of the formation of organic-carbon-rich sapropels.

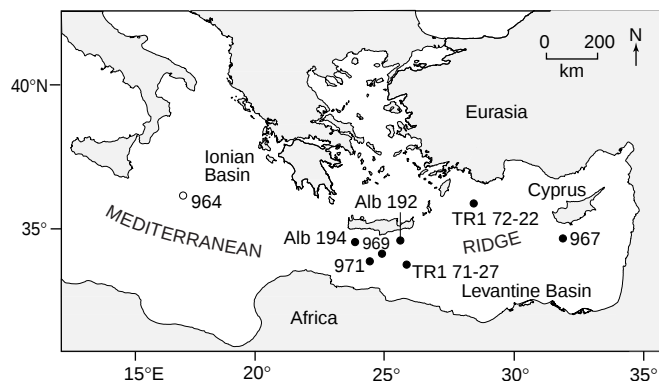
To analyse the composition of individual laminae, we utilized electron-microscope techniques involving back-scattered electron imagery (BSEI) of sediment prepared by fluid-dispersive resin-embedding in order to preserve the delicate microfabrics<sup>8</sup>. Having used BSEI to define the laminae (Fig. 2), we then analysed floral assemblages by using topographic electron-microscope imagery of sub-samples taken from counterparts of laminae (Fig. 2) and conventional diatom micropalaeontological preparations sampled, in surgical style, from groups of laminae. The sediment consists of

diatomaceous laminae intercalated with subordinate, variably thick layers of mineralogenic detritus which we interpret to represent either exudates of the Napoli mud volcano or material shed from its flanks in low-density gravity flows. BSEI analysis of a 50-cm section of the sapropel, containing 976 laminae/layers, reveals a regularly recurring couplet of diatomaceous laminae; this couplet consists of a pure, near-monospecific diatom lamina alternating with a lamina composed of a mixed diatom assemblage together with minor coccolith debris (Fig. 2). Most (86%) of the former laminae comprise mat-forming rhizosolenid diatoms with 14% composed of the chain-forming *Hemiaulus hauckii* (Fig. 2) and, occasionally, *Ethmodiscus* sp. Conventional diatom preparations reveal that the rhizosolenids are dominantly *Pseudosolenia calcar-avis* with smaller amounts of *Rhizosolenia styliformis* and rare *R. bergonii* and *R. setigera*<sup>8</sup>. Rhizosolenid diatoms comprise 46% of the total valves counted from 17 samples (see Supplementary Information)<sup>8,9</sup>. The laminae containing mixed diatom assemblages contain mainly *Thalassionema frauenfeldii*, with smaller abundances of *T. nitzschoides*, *Bacteriastrum* spp. and *Chaetoceros* spp. resting spores and few rhizosolenids. Another sample of S5 from ODP Site 967 south of Eratosthenes seamount (Fig. 1) has a similar diatom assemblage (see Supplementary Information) although the frustules have undergone considerable silica dissolution<sup>9</sup>. Similar diatom floras were previously identified in sapropels S4 and S5 from the eastern Mediterranean using conventional micropalaeontological methods<sup>10</sup> (Fig. 1). An early Pleistocene diatomaceous sapropel from ODP Site 969 located near Site 971 on the Mediterranean ridge (Fig. 1) also contains abundant rhizosolenids<sup>9</sup> (see Supplementary Information).

We assume an annual periodicity for the lamina couplets in S5 at Site 971 on the basis of the regular recurrence of the two lamina types, which points to two distinct seasonal episodes of flux. The seasonal flux cycle may be reconstructed by consideration of the ecology of the diatoms involved. The most common diatom within the mixed assemblage laminae, *T. frauenfeldii*, has been recorded as the dominant diatom in the winter/early-spring period in the modern eastern Mediterranean<sup>11,12</sup>, and *T. nitzschoides*, *Bacteriastrum* spp. and *Chaetoceros* spp. resting spores are also common<sup>12</sup>. The intact nature of many of the diatoms in these laminae suggests the rapid flocculation and sinking associated with blooms; such processes are well known, both from water-column observation<sup>13</sup> and experimental mesocosm study, and from BSEI studies of sedimentary laminae<sup>14</sup>. We may suppose, therefore, that the mixed-assemblage lamina represents the fallout from production that occurred during the winter-spring period.

The interpretation of the rhizosolenid laminae is not so straightforward. In fact, both *P. calcar-avis* and *H. hauckii* are characteristic of the modern Mediterranean, but typically occur in low abundances through nutrient-poor conditions during the months of summer stratification (May–October)<sup>12,15</sup>. Indeed, large rhizosolenid diatoms and *Hemiaulus* spp. have been generally regarded as slow-growing and characteristic of the sparse flora of oligotrophic waters<sup>15</sup>. Laboratory and field studies have demonstrated that both rhizosolenids and *Hemiaulus* spp. have specific adaptations for stratified oligotrophic waters including symbiosis with nitrogen-fixing bacteria<sup>16</sup>. Also, rhizosolenid mats are capable of buoyancy regulation, allowing them to sink to withdraw nitrogen from the nutricline and rise to higher levels within the euphotic zone to photosynthesize<sup>5</sup>. A further adaptation to stratified waters in several species of large or colonial diatoms, including rhizosolenids, enables them to bloom in low-light conditions near the lowermost part of the euphotic zone<sup>17</sup>.

Given their characteristic low abundance in stratified oligotrophic waters, how then do these diatoms become concentrated to form monospecific laminae representing massive flux? The excellent preservation of the delicate rhizosolenid frustules attests to rapid deposition. We therefore require a mechanism to account



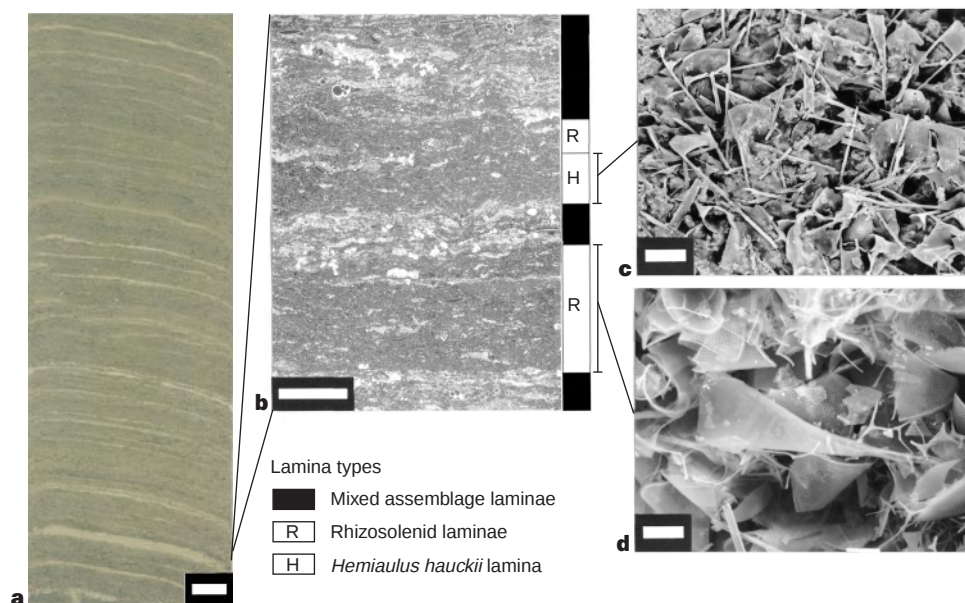
**Figure 1** Summary map of the eastern Mediterranean. The location of the relevant ODP Leg 160 sites and other cores containing diatomaceous sapropels<sup>10</sup> are indicated. Filled symbols indicate sites that contain diatomaceous sapropels, including ODP Sites 967, 969 and 971.

for both the concentration of rhizosolenids and their rapid mass-sinking. Mass concentration<sup>6</sup> and sinking<sup>18</sup> of rhizosolenid diatom mats monitored in the equatorial Pacific, and vast Neogene *Thalassiothrix* diatom mat deposits from the same region<sup>7</sup>, have been related to the action of tropical instability waves on the front between the South Equatorial Current and the North Equatorial Counter Current<sup>6</sup>. Such a model is unlikely to apply in the eastern Mediterranean as no frontal systems analogous to the equatorial Pacific operate across the area of occurrence of rhizosolenid deposits (Fig. 1).

An alternative mechanism for mass sinking within the annual cycle that does not require the action of fronts may be surmised from a recent synthesis of the timing of massive rhizosolenid diatom flux (which generally occurs in the autumn–winter<sup>19</sup>) together with recent studies of Gulf of California sediments. The seasonal flux pattern of diatomaceous laminae recorded in S5 resembles that documented by sediment-trap data<sup>20,21</sup> and BSEI analysis<sup>22</sup> of laminated Holocene sediments for the Gulf of California. There, a spring bloom is recorded by dominant flux of *Chaetoceros* spp. resting spores. During the period of summer stratification and minimal biogenic flux in the Gulf, diatom floras include both *P. calcar-avis* and *H. hauckii*<sup>21</sup> (as in the Mediterranean). Autumn/early-winter deposition, coinciding with breakdown of stratification, is represented by mat-forming diatoms including rhizosolenids, *Thalassiothrix* spp. mats as well as other colonial and large solitary diatoms documented by Goldman<sup>17</sup> as growing at depth in low-light conditions. Such diatoms, especially rhizosolenids, are very sensitive to water-column agitation<sup>23</sup>, and sediment massively when mixing breaks down summer stratification<sup>19,23</sup>. We therefore attribute the rhizosolenid/*H. hauckii* laminae occurring in S5 to an autumn/early-winter flux caused by the onset of mixing and breakdown of stratification in the Mediterranean. Thus the annual cycle of flux in the Mediterranean at the time of sapropel formation was dominated by fallout from winter–spring blooms, followed by protracted growth over the summer period of stratification which was terminated by massive sedimentation of diatoms at the onset of autumn/early-winter mixing (Fig. 3).

We now show how our reconstruction of the annual cycle of production and flux reconciles apparently divergent existing models of sapropel formation, and thus confirms earlier speculation of the role of diatoms by Sancetta<sup>24</sup>. There is general agreement that sapropel formation occurred during periods of enhanced freshwater runoff to the eastern Mediterranean. This occurred at times of precession minima through intensification of the summer monsoon resulting in increased discharge from the Nile<sup>25</sup> and probably also from increased influence of Atlantic-derived depressions causing elevated river discharge along the northern borderlands of the eastern Mediterranean<sup>1</sup>. This increased freshwater input led to enhanced stratification of surface waters. Nutrients imported through runoff to the eastern Mediterranean would rapidly have been recycled and trapped in a nutricline-associated halocline. These conditions were ideally suited to diatoms adapted to stratified conditions. There is evidence of the common existence of a deep chlorophyll maximum during periods of sapropel formation, both from the abundance of neogloboquadrinid foraminifera<sup>1</sup> and the coccolithophorid *Florisphaera profunda*<sup>2</sup>. On the basis of the evidence presented above, it is now clear that the dominant agent of production in this deep chlorophyll maximum and the food source for the neogloboquadrinids were diatoms. The relative contributions of production at depth (probably 100–130 m; refs 1, 2, 17) in this chlorophyll maximum and of production near the surface through vertical migration of rhizosolenid mats is uncertain. Winter mixing would have broken down stratification<sup>26,27</sup> and caused the mass sinking of rhizosolenid and *H. hauckii* diatoms which had been growing progressively through the summer. The mixing was sufficient to introduce adequate nutrients to the surface to generate conventional, near-surface blooms during winter–early spring recorded by the lamina of mixed diatom species.

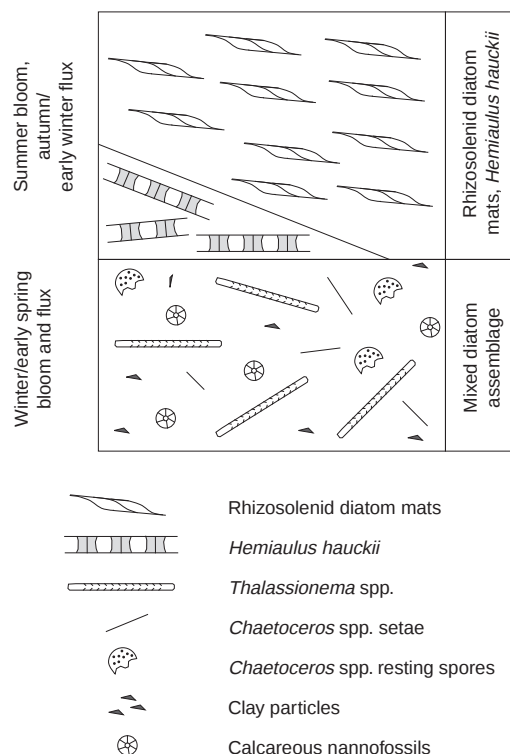
To quantify palaeoproductivity and to assess the contribution from diatoms, we analysed organic carbon and biogenic silicon ( $\text{Si}_{\text{bio}}$ ) contents (Table 1). We combined these data with sedimentation rates based on the annual occurrence of lamina couplets (average thickness  $550 \mu\text{m yr}^{-1}$ ) and measurements of dry bulk density to produce flux data. Organic carbon and biogenic silicon



**Figure 2** Composition of the laminated sapropel. **a**, Core photograph of the laminated diatom ooze sapropel (S5), core interval 160-971C-2H-3, 81–110 cm (scale bar, 1 cm). **b**, Back-scattered electron image (BSEI) of a polished thin section of fluid-displacive resin-embedded sediment showing rhizosolenid, mixed assemblage and *H. hauckii* laminae, core interval 160-971C-2H-3, 107–110 cm. The darker-tone rhizosolenid laminae are much more porous than the

mixed-assemblage laminae as they contain a greater proportion of carbon-based resin with a low backscatter coefficient (scale bar, 1 mm). **c**, BSEI of a well-preserved rhizosolenid-rich lamina (sample 160-971C-2H-3, 126–129 cm) containing *P. calcar-avis* process (centre) and girdle bands that make up the bulk of the lamina thickness (scale bar, 10  $\mu\text{m}$ ). **d**, BSEI of a *H. hauckii* sub-lamina (sample 160-971C-2H-3, 106.5–111 cm) (scale bar, 20  $\mu\text{m}$ ).





**Figure 3** Annual cycle of production and flux forming sapropel S5. Figure shows a schematic representation of the annual cycle of production and flux contained in the two-component varve common within the laminated diatom ooze of sapropel S5, Site 971, hole C. The *Thalassionema* spp. component of the mixed-assemblage laminae comprises *T. frauenfeldii* with subsidiary quantities of *T. nitzschoides*. The rhizosolenid diatom laminae are dominated by *P. calcar-avis*.

accumulation rates range from 4.6 to 9.2 g C m<sup>-2</sup> yr<sup>-1</sup> and 28 to 95 g Si<sub>bio</sub> m<sup>-2</sup> yr<sup>-1</sup> respectively, which are comparable to modern high-productivity oceanic sites. An indication of the diatom contribution to export production and the total organic carbon content may be had by comparing an original diatom Si:C ratio (as measured by Brzezinski<sup>28</sup>) of 0.14 with the present average value in S5 (from Table 1) of 2.9. This would give an export production (due to diatoms) in the range 100–200 g C m<sup>-2</sup> yr<sup>-1</sup> (we assume minimal opal dissolution as evidenced by the excellent diatom preservation). The notional 20 times decrease in organic carbon seems a sufficiently realistic degradation within an anoxic environment (compare Black Sea organic carbon preservation of 5%; ref. 29) to suggest that the diatom contribution may account for the entire organic-carbon content of the sapropel. Sapropels that lack opal and have organic-carbon contents up to 30% may be regarded as remnant deposits in which the organic carbon has been concentrated by opal dissolution.

The crucial role of diatoms in sapropel formation was inferred by van Os *et al.*<sup>26</sup>, who cite the late Pliocene, Cretan and Sicilian rhythmic diatomites, which are coeval with sapropels in the Mediterranean basin, as evidence of original diatom content. The intermittent preservation of opal in sapropels appears to be more common in the most tectonically active areas, where the periodic formation of topographic depressions may foster the development of anoxic basins (for example, the Mediterranean ridge). Our investigations of sapropels from local topographic highs such as Site 964 on the Pisano plateau (Fig. 1) show no opal preservation<sup>9</sup>.

The implications of this study apply to earlier geological times and are relevant outside the Mediterranean basin. There has been a tendency to interpret evidence of high diatom production as upwelling-related, both in earlier studies of the Mediterranean

**Table 1** Opal and carbon determinations

| Dataset | Sample number | Lamina number(s) comprising each sample* | Cont.† | Si <sub>bio</sub> (%)‡ | Opal (%)§ | TOC (%) |
|---------|---------------|------------------------------------------|--------|------------------------|-----------|---------|
| 1       | 1             | 79–80                                    | P      | 10.31                  | 24.74     | NA      |
|         | 2             | 72–76                                    | P      | 13.83                  | 33.18     | NA      |
|         | 3             | 67–71                                    | P      | 11.55                  | 27.72     | NA      |
|         | 4             | 62–66                                    | P      | 9.90                   | 23.77     | NA      |
|         | 5             | 62–66                                    | P      | 8.15                   | 19.57     | NA      |
|         | 6             | 58–66                                    | P      | 10.82                  | 25.96     | NA      |
|         | 7             | 50–59                                    | P      | 9.43                   | 22.64     | NA      |
|         | 8             | 42–50                                    | P      | 12.43                  | 29.84     | NA      |
|         | 9             | 38–41                                    |        | 16.24                  | 38.97     | NA      |
|         | 10            | 34–36                                    |        | 12.67                  | 30.40     | NA      |
|         | 11            | 31–36                                    |        | 15.65                  | 37.57     | NA      |
|         | 12            | 25–31                                    |        | 14.87                  | 35.68     | NA      |
|         | 13            | 22–27                                    |        | 24.15                  | 57.96     | NA      |
|         | 14            | 15–21                                    |        | 13.13                  | 31.50     | NA      |
|         | 15            | 9–14                                     |        | 13.98                  | 33.56     | NA      |
|         | 16            | 6–10                                     |        | 12.90                  | 30.97     | NA      |
|         | 17            | 1–5                                      | P      | 7.23                   | 17.35     | NA      |
| 2       | 24            | 188–242                                  |        | 16.14                  | 38.74     | 3.95    |
|         | 25            | 172–187                                  | P      | 13.07                  | 31.37     | 3.83    |
|         | 26            | 163–171                                  |        | 15.58                  | 37.40     | 3.61    |
|         | 27            | 156–162                                  | D      | 6.48                   | 15.55     | 2.75    |
|         | 28            | 155                                      | P      | 13.70                  | 32.89     | 5.60    |
|         | 29            | 127–154                                  | P      | 12.50                  | 30.01     | 4.73    |
|         | 30            | 109–126                                  | D      | 8.77                   | 21.04     | 5.24    |
|         | 31            | 82–108                                   | P      | 13.26                  | 31.82     | 5.36    |
|         | 32            | 79–80                                    | P      | 11.61                  | 27.85     | 3.44    |
|         | 33            | 28–77                                    | P      | 13.12                  | 31.49     | 4.95    |
|         | 34            | 5–27                                     |        | 15.65                  | 37.56     | 4.43    |

\* Lamina subdivision was on the basis of changes in diatom composition. Sample intervals were determined by the ease with which individual laminae could be identified in core section.

† Contamination: P, possible contamination by terrigenous laminae; D, definite contamination by terrigenous laminae.

‡ Biogenic silicon.

§ Opal (SiO<sub>2</sub>·0.4H<sub>2</sub>O) = 2.4 × Si<sub>bio</sub>(Si) (ref. 32).

|| Total organic carbon. The biogenic silicon determination was undertaken using a wet-alkaline extraction technique, organic carbon was measured by combustion using a Carbo Erba elemental analyser. NA, not available.

sapropels<sup>10</sup> and in the interpretation of Palaeogene and Cretaceous diatomites<sup>30</sup>, yet examination of the diatoms involved suggests an alternative explanation. In fact, both rhizosolenids and the genus *Hemiaulus* were diverse and common components of Palaeogene and Late Cretaceous diatom floras<sup>30</sup>. It follows that these diatoms grew under conditions of sustained seasonal stratification, rather than in response to upwelling. Furthermore, the sapropels resemble the rhythmic black shales of the Cretaceous, which also typically occur at frequencies associated with orbital precession or eccentricity, and which have been related to monsoonal periods of enhanced runoff carried by rivers<sup>31</sup>. Although diatoms are rarely preserved in sediments of Cretaceous age, it may be that, as with the Mediterranean sapropels, diatom production contributed to the organic-rich nature of these sediments and to the genesis of other petroleum source rocks. □

Received 17 August 1998; accepted 7 January 1999.

- Rohling, E. J. Review and new aspects concerning formation of the Mediterranean sapropels. *Mar. Geol.* **122**, 1–28 (1994).
- Castradori, D. Calcareous nannofossils and the origin of eastern Mediterranean sapropels. *Paleoceanography* **8**, 459–471 (1993).
- Calvert, S. E., Nielsen, B. & Fontugne, M. R. Evidence from nitrogen isotope ratios for enhanced productivity during formation of eastern Mediterranean sapropels. *Nature* **359**, 223–225 (1992).
- De Lange, G. J. *et al.* Composition of anoxic hypersaline brines in the Tyro and Bannock Basins, eastern Mediterranean. *Mar. Chem.* **31**, 63–88 (1990).
- Villareal, T. A., Altabet, M. A. & Culver-Rymasz, K. Nitrogen transport by vertically migrating diatom mats in the North Pacific Ocean. *Nature* **363**, 709–712 (1993).
- Yoder, J. A., Ackleson, S., Barber, R. T., Flamant, P. & Balch, W. A. A line in the Sea. *Nature* **371**, 689–692 (1994).
- Kemp, A. E. S. & Baldauf, J. G. Vast Neogene laminated diatom mat deposits from the eastern equatorial Pacific Ocean. *Nature* **362**, 141–143 (1993).
- Pearce, R. B. *et al.* A lamina-scale, SEM-based study of a diatom-ooze sapropel from the Eastern Mediterranean (ODP Leg 160). *Proc. ODP Sci. Res.* **160**, 333–348 (1998).
- Kemp, A. E. S., Pearce, R. B., Pike, J. & Marshall, J. E. A. Microfabric and microcompositional studies of Pliocene and Quaternary sapropels from the eastern Mediterranean. *Proc. ODP Sci. Res.* **160**, 349–364 (1998).

10. Schrader, H. J. & Matherne, A. Sapropel formation in the eastern Mediterranean sea: evidence from preserved opal assemblages. *Micropalaeontology* **27**, 191–203 (1981).
11. Pagou, K. & Gotsis-Skretas, O. A comparative study of phytoplankton in S. Aegean, Levantine and Ionian Seas during March–April. *Thalassographica* **13**, 13–18 (1990).
12. Ignatiades, L., Georgopoulos, D. & Karydis, M. Description of the phytoplanktonic community of the oligotrophic waters of the SE Aegean Sea (Mediterranean). *Mar. Ecol.* **16**, 13–26 (1995).
13. Alldredge, A. L. & Gotschalk, C. C. Direct observations of the mass flocculation of diatom blooms: characteristics, settling velocities and formation of diatom aggregates. *Deep-Sea Res.* **36**, 159–171 (1989).
14. Grimm, K. A., Lange, C. B. & Gill, A. S. Self-sedimentation of phytoplankton blooms in the geologic record. *Sedim. Geol.* **100**, 151–161 (1997).
15. Guillard, R. R. L. & Kilham, P. In *The Biology of Diatoms* (ed. Werner, D.) 372–469 (Botanical Monogr. 13, Univ. California Press, Berkeley, 1977).
16. Villareal, T. A. Nitrogen-fixation by the cyanobacterial symbiont of the diatom genus *Hemiaulus*. *Mar. Ecol. Prog. Ser.* **76**, 201–204 (1991).
17. Goldman, J. C. Potential role of large oceanic diatoms in new primary production. *Deep-Sea Res.* **140**, 159–168 (1993).
18. Smith, C. R. *et al.* Phytodetritus at the abyssal seafloor across 10° of latitude in the central equatorial Pacific. *Deep-Sea Res.* **43**, 1309–1338 (1996).
19. Sancetta, C., Villareal, T. & Falkowski, P. Massive fluxes of Rhizosolenid diatoms: a common occurrence? *Limnol. Oceanogr.* **36**, 1452–1457 (1991).
20. Thunell, R., Pride, C., Tappa, E. & Muller-Karger, F. Varve formation in the Gulf of California: Insights from time series sediment trap sampling and remote sensing. *Quat. Sci. Rev.* **12**, 451–464 (1993).
21. Sancetta, C. Diatoms in the Gulf of California: Seasonal flux patterns and the sediment record for the last 15,000 years. *Paleoceanography* **10**, 67–84 (1995).
22. Pike, J. & Kemp, A. E. S. Early Holocene decadal-scale ocean variability recorded in Gulf of California laminated sediments. *Paleoceanography* **12**, 227–238 (1997).
23. Richardson, T. L., Ciotti, A. M., Cullen, J. J. & Villareal, T. A. Physiological and optical properties of *Rhizosolenia formosa* (Bacillariophyceae) in the context of open-ocean vertical migration. *J. Phycol.* **32**, 741–757 (1996).
24. Sancetta, C. Mediterranean sapropels: Seasonal stratification yields high production and carbon flux. *Paleoceanography* **9**, 195–196 (1994).
25. Rossignol-Strick, M., Nesteroff, V., Olive, P. & Vergnaud-Grazzini, C. After the deluge; Mediterranean stagnation and sapropel formation. *Nature* **295**, 105–110 (1982).
26. Van Os, B. J. H., Lourens, L. J., Hilgen, F. J., De Lange, G. J. & Beaufort, L. The formation of Pliocene sapropels and carbonate cycles in the Mediterranean: diagenesis, dilution and productivity. *Paleoceanography* **9**, 601–617 (1994).
27. Tang, C. M. & Stott, L. D. Seasonal salinity changes during Mediterranean sapropel deposition 9,000 years B.P.: evidence from isotopic analyses of individual planktonic foraminifera. *Paleoceanography* **8**, 473–493 (1993).
28. Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *J. Phycol.* **21**, 347–357 (1985).
29. Arthur, M. A. *et al.* Varve calibrated records of carbonate and organic carbon accumulation over the last 2000 years in the Black Sea. *Glob. Biogeochem. Cycles* **8**, 195–217 (1994).
30. Dell'Agnes, D. J. & Clark, D. L. Siliceous microfossils from the warm late Cretaceous and early Cenozoic Arctic Ocean. *J. Paleontol.* **68**, 31–47 (1994).
31. Jacobs, D. K. & Sahagian, D. L. In *Sequence Stratigraphy and Depositional Response to Eustatic, Tectonic and Climatic Forcing* (ed. Haq, B. U.) 329–366 (Kluwer, Dordrecht, 1995).
32. Mortlock, R. A. & Froelich, P. N. A simple method for the rapid determination of biogenic opal in marine sediments. *Deep-Sea Res.* **36**, 1415–1426 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We acknowledge funding from the NERC ODP research grant and research fellowship programme.

Correspondence and requests for materials should be addressed to A.E.S.K. (e-mail: [aesk@soc.soton.ac.uk](mailto:aesk@soc.soton.ac.uk)).

## Low-temperature thermal generation of hydrocarbon gases in shallow shales

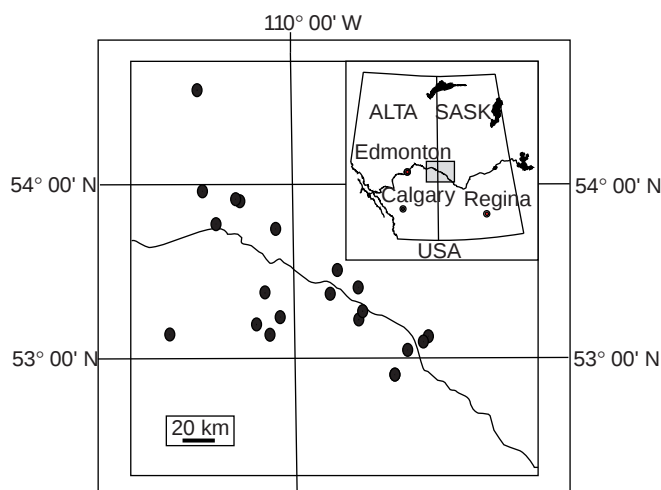
Devon Rowe & Atis Muehlenbachs

Department of Earth and Atmospheric Sciences, University of Alberta, 1-26 ESB, Edmonton AB T6G 2E3, Canada

The thermal generation of hydrocarbon gases in sedimentary systems is generally thought to be an exclusively high-temperature process<sup>1</sup>, although previous work has indicated that significant generation may take place at burial temperatures as low as 65°C (ref. 2). Here we present the carbon-isotope signatures of gases from the Western Canadian sedimentary basin. The isotope data show that low-temperature thermal generation of non-methane hydrocarbons occurs at temperatures lower than 62°C and possibly as low as 20°C. We can distinguish between gases of shallow<sup>3,4</sup> and deep origin by using the carbon-isotope compositions of the non-methane components (ethane, propane and butane); the shallow gases have isotopic compositions consistent

with a thermal origin, whereas the deep gases—although of abiogenic origin—bear isotopic signatures that have been biologically altered. These findings have allowed the development of a successful tool for detecting the source of leaking gases around oil wells in this basin, enabling cheaper and more effective remediation.

The Upper Cretaceous Colorado Group of the Western Canadian sedimentary basin (WCSB) is a sequence of marine shales that vary in total organic content from 4 wt% to 12 wt% (ref. 5). Beneath the Colorado Group lies the Mannville Group, which consists of heavy-oil bearing sands with interbedded shales, siltstones and fine-grained sandstones. The predominant gas within these two Groups is methane (CH<sub>4</sub>), with small amounts of ethane (C<sub>2</sub>H<sub>6</sub>), propane (C<sub>3</sub>H<sub>8</sub>) and butane (C<sub>4</sub>H<sub>10</sub>) being present (Table 1). Major erosional events have removed sediments from above the Colorado and Mannville Groups since their maximum burial period; hydrocarbon generation would probably have occurred before these erosional events. Present-day conditions (depth and temperature), therefore, are not representative of the true generation conditions. Optical properties (vitrinite reflectance) of buried coal can be used to gauge the thermal maturity of the organic matter and hence to estimate maximum burial depths. Published vitrinite-reflectance values ranging from 0.3% to 0.4% indicate that the Mannville Group is thermally immature<sup>6,7</sup>. The Colorado Group does not contain coal for direct measurement of vitrinite reflectance, but an estimate can be made from established vitrinite-reflectance depth gradients in the area and from a measured value for vitrinite reflectance from within the Mannville Group (one sample from 556 m in our study area has a value for vitrinite reflectance of 0.39%). This data point, combined with the gradients published in ref. 6, results in estimates for vitrinite reflectance of 0.25–0.3% for the Colorado Group. These values reflect the combined effects of the present-day burial and the extra sediments that have been eroded.



**Figure 1** Study area in the Western Canadian sedimentary basin, Canada. The location of 20 sites at which shallow natural gases were collected. At each location, drilling mud was collected as it exited the drill stem at the surface. Each well was drilled overbalanced, and sampling intervals were at roughly every 50 m. The mud (which contained fresh water and cuttings from the drill bit) was collected in 1-litre plastic containers as it exited the drill pipe (that is, at the shale shaker). A headspace was left in each container for gases held within the fluid to occupy. The 'mud gases' were heated and agitated to liberate them from solution before extracting the headspace gas for analysis of bulk gas and carbon-isotope composition (Table 1). The  $\delta^{13}\text{C}$  values (relative to Pee Dee Belemnite (PDB) standard) of the hydrocarbons were determined using a Finnigan MAT 252 (ConFlo II) continuous-flow isotope-ratio mass spectrometer, with precisions of  $\pm 0.2\%$  for methane and ethane and  $\pm 0.3\%$  for propane and butane. Data from the well indicated with an arrow are shown in Fig. 2 and Table 1.

**Table 1 Composition of Mannville Group gases**

| Sample depth<br>(m below surface) | Methane                            |        | Ethane                             |        | Propane                            |        | <i>i</i> -Butane                   |        | <i>n</i> -Butane                   |        |
|-----------------------------------|------------------------------------|--------|------------------------------------|--------|------------------------------------|--------|------------------------------------|--------|------------------------------------|--------|
|                                   | $\delta^{13}\text{C}_{\text{C}_1}$ | Vol. % | $\delta^{13}\text{C}_{\text{C}_2}$ | Vol. % | $\delta^{13}\text{C}_{\text{C}_3}$ | Vol. % | $\delta^{13}\text{C}_{\text{C}_4}$ | Vol. % | $\delta^{13}\text{C}_{\text{C}_4}$ | Vol. % |
| 195                               | -67.7                              | 99.6   | -54.1                              | 0.3    | -44.9                              | 0.05   | -35.7                              | 0.02   | ND*                                | 0.01   |
| 236                               | -67.1                              | 99.4   | -52.0                              | 0.4    | -43.2                              | 0.1    | ND*                                | 0.03   | ND*                                | 0.01   |
| 305                               | -65.0                              | 98.7   | -51.4                              | 1.0    | -41.5                              | 0.2    | -33.3                              | 0.06   | -35.3                              | 0.03   |
| 381                               | -64.4                              | 98.4   | -48.0                              | 1.2    | -42.8                              | 0.2    | -32.8                              | 0.06   | -36.5                              | 0.04   |
| 404                               | -63.2                              | 98.4   | -47.7                              | 1.3    | -41.2                              | 0.2    | -34.0                              | 0.05   | -38.0                              | 0.03   |
| 450                               | -63.0                              | 98.5   | -44.8                              | 1.3    | -41.0                              | 0.2    | -32.1                              | 0.02   | -36.1                              | 0.03   |
| 490†                              | -63.0                              | 98.7   | -43.6                              | 1.0    | -34.9                              | 0.2    | -32.4                              | 0.04   | -32.4                              | 0.03   |
| 567†                              | -60.9                              | 98.0   | -30.1                              | 1.7    | -26.4                              | 0.2    | -28.5                              | 0.05   | -31.6                              | 0.5    |
| 610                               | -57.8                              | 92.4   | -30.4                              | 1.6    | -23.4                              | 0.9    | -27.6                              | 1.0    | -25.2                              | 0.8    |
| 650                               | -57.0                              | 93.8   | -31.6                              | 4.0    | -23.8                              | 0.6    | -23.8                              | 0.5    | -24.8                              | 0.5    |

\* Not detectable.

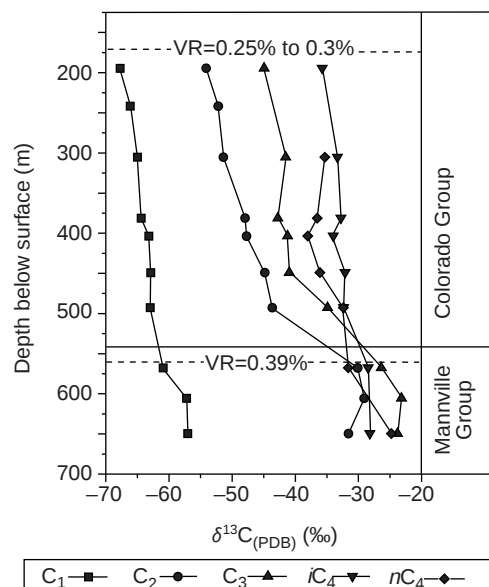
† The Colorado–Mannville Group boundary is at a depth of 541 m, identifiable by the change in isotopic composition of the  $\text{C}_{2+}$  components.

By extrapolating the vitrinite-reflectance gradients in the basin<sup>7</sup>, we estimate the thickness of the eroded section to be 1,500 m. However, geomorphological evidence indicates that only 450 m may have been removed<sup>8</sup>. On the basis of the present-day depths (200–550 m) and palaeothermal gradients ( $30^\circ\text{C km}^{-1}$ )<sup>6</sup>, we estimate that burial temperatures in the Colorado Group would have been as low as  $20^\circ\text{C}$ , and would not have exceeded  $62^\circ\text{C}$ .

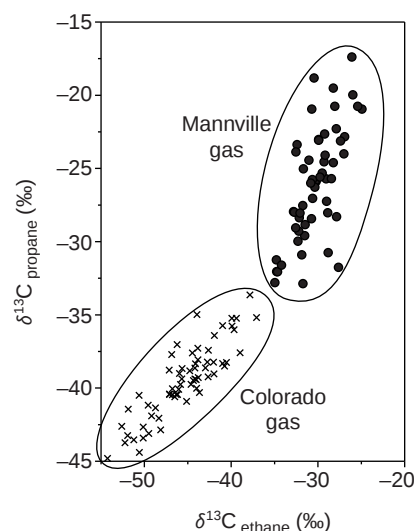
We analysed isotope ratios of methane ( $\text{C}_1$ ), ethane ( $\text{C}_2$ ), propane ( $\text{C}_3$ ) and butane ( $\text{C}_4$ ) components from gases collected at 50-m depth intervals during the drilling of 20 oil wells across the study area (Fig. 1). The molecular compositions vary from 90 to 99 vol.% methane, with  $\text{C}_{2+}$  components (ethane, propane and butane) accounting for up to 5 vol.%. In most cases, gases from the Colorado and Mannville Groups have such similar bulk compositions that they cannot be distinguished (Table 1). The  $\text{C}_{2+}$  carbon-isotope ratios, however, provide an unambiguous distinction.

The  $\delta^{13}\text{C}$  values are correlated with stratigraphy, with significant  $^{13}\text{C}$  enrichment occurring in the Mannville Group gases (Fig. 2). We relied less on the isotopic composition of the methanes and

instead used the  $\delta^{13}\text{C}_{\text{C}_{2+}}$  values. (The low  $\delta^{13}\text{C}_{\text{C}_1}$  and high vol.%( $\text{C}_1$ ) (Table 1) indicate that the methanes may be derived bacterially, perhaps through different mechanisms and from multiple sources. Hence,  $\delta^{13}\text{C}_{\text{C}_1}$  values cannot be unambiguously related to any particular source characteristics<sup>9</sup>). The  $\delta^{13}\text{C}_{\text{C}_{2+}}$  values differ between the Colorado and Mannville Groups, and suggest that the Colorado ( $\delta^{13}\text{C}_{\text{C}_2}$ , 63‰ to -32‰;  $\delta^{13}\text{C}_{\text{C}_3}$ , -42‰ to -23‰;  $\delta^{13}\text{C}_{\text{C}_4}$ , -37‰ to -22‰) and Mannville ( $\delta^{13}\text{C}_{\text{C}_2}$ , -30‰ to -22‰;  $\delta^{13}\text{C}_{\text{C}_3}$ , -27‰ to -15‰;  $\delta^{13}\text{C}_{\text{C}_4}$ , -34‰ to -20‰) groups contain two very different gas families. The presence of  $\text{C}_{2+}$  components indicates that gases in both Groups are derived abiologically, but the Colorado Group gases show three unique trends: first,  $\delta^{13}\text{C}$  values of shallow  $\text{C}_{2+}$  gases increase with depth; second,  $\delta^{13}\text{C}$  values of  $\text{C}_{3+}$  components at a given depth increase with the number of carbon atoms; and third, the difference in isotopic composition between hydrocarbon components tends to increase at shallow depths. These trends are diagnostic of thermal gases<sup>10–12</sup> and are characteristic of Colorado Group  $\text{C}_{2+}$  gases from each well sampled. These thermal characteristics are not, however, found in the Mannville Group  $\text{C}_{2+}$



**Figure 2** 'Mud-gas' isotopic depth profile or 'fingerprint'. The vitrinite reflectance (VR) measured from one coal sample of the Mannville Group is indicated, as is an estimated lower limit for vitrinite reflectance for the Colorado Group (calculated using the measured data point and the known VR gradient in the area). Across the Colorado–Mannville geological boundary (located at 541 m), there is a dramatic change in isotopic compositions (a change of  $\sim 15\%$  for  $\text{C}_2$  and  $\sim 10\%$  for  $\text{C}_3$ ). Mannville Group gases have a characteristic isotopic reversal between  $\text{C}_3$  and  $\text{C}_4$  components.



**Figure 3**  $\delta^{13}\text{C}_{\text{propane}}$  versus  $\delta^{13}\text{C}_{\text{ethane}}$  values for gases from the Colorado and Mannville Groups. These gases plot in two distinct regions. The enriched  $\delta^{13}\text{C}_{\text{propane}}$  values of the Mannville gases are due to microbial activity<sup>4,13</sup>. The Colorado gases are so immature (very negative  $\delta^{13}\text{C}$  values) that they cannot be compared to results obtained from published kinetic models. The isotopic separations between the  $\text{C}_3$  and  $\text{C}_2$  components however, are consistent with the separations predicted by a 'quasi-equilibrium' approach<sup>10</sup>.

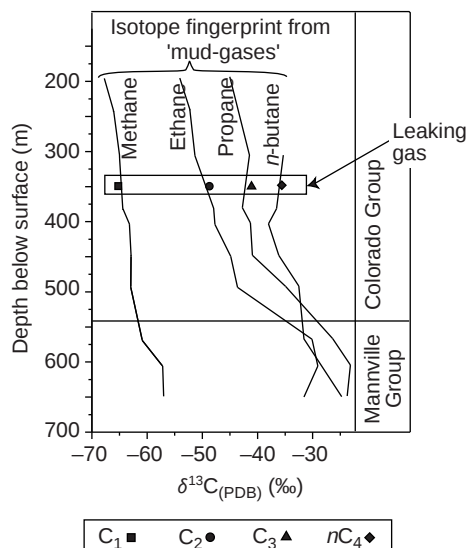


gases; in particular, the  $\delta^{13}\text{C}_{\text{C}_4}$  values are commonly less negative than the  $\delta^{13}\text{C}_{\text{C}_3}$  values, a result of microbial alteration<sup>13</sup>.

In contrast to the  $\text{C}_{2+}$  gases in the Michigan basin<sup>14</sup>, the  $\text{C}_{2+}$  components of the Colorado gases in the WCSB could not have migrated from deeper formations. If the gases in the Mannville formations had migrated vertically into the Colorado shales, isotope ratios of  $\text{C}_{2+}$  in both groups would be similar, but this is not the case (Figs 2, 3). Isotopic reversals between  $\text{C}_3$  and  $\text{C}_4$  components of Mannville gases (for example,  $\delta^{13}\text{C}_{\text{C}_4} < \delta^{13}\text{C}_{\text{C}_3}$ ; Fig. 2) do not occur in the Colorado shale gases<sup>3</sup>.

Ratios of  $^{13}\text{C}/^{12}\text{C}$  for  $\text{C}_1$  to  $\text{C}_4$  hydrocarbons have been used several times to investigate gas-generation mechanisms. Several conceptual models have been developed that incorporate kinetic<sup>11,12,15–18</sup>, equilibrium<sup>10</sup> or catalytic processes<sup>17,19</sup>. For example, given some knowledge of source conditions,  $\delta^{13}\text{C}$  values of economic gas deposits can often be accurately predicted from kinetic models<sup>16,18</sup>. However, these models are usually restricted to cases in which the source material being considered is much more mature than the Colorado Group (that is, cases in which vitrinite reflectance is more than 0.4%) and the formation temperatures exceed 80 °C (that is, peak stages of hydrocarbon generation). At the low temperature/maturity limit, trends of the kinetic models are still consistent with the large isotopic separations and very negative  $\delta^{13}\text{C}_{\text{C}_{2+}}$  values seen in the Colorado gases. In the shallow WCSB, however, extrapolation beyond the limits of the kinetic models (for example, to the source conditions of the Colorado Group) seems invalid. In the 'quasi-equilibrium' approach<sup>10</sup>, there are no such restrictions on source immaturity. The equilibrium isotopic separations predicted for gases generated from very immature source materials (that is, materials with vitrinite reflectance values of  $>0.4\%$ ; see level of organic metamorphism, LOM,  $< 6$  on Figure 2 of ref. 3) in fact fit the observed Colorado-gas  $\delta^{13}\text{C}$  values very well. Transition-metal<sup>17</sup> or mineral<sup>19</sup> catalysis may control hydrocarbon generation, and this may explain the formation of natural gas in our study area, which shows approximate low-temperature isotopic equilibrium.

The discovery of natural gas in the Colorado and Mannville



**Figure 4** Isotopic fingerprint used to estimate the source of a leaking gas. Isotope ratios of a migrating gas from a leading well are compared to the isotopic fingerprint for that well. By matching the isotopic separation of the  $\text{C}_1$  to  $\text{C}_4$  components with the fingerprint, the source depth of the leaking gas can be determined, being roughly 350 m in this case (fingerprint does not include  $i\text{C}_4$ , as no  $i\text{C}_4$  was detectable in leaking gas). Accurate determination of the source of leaking gases has greatly increased remediation success in the WCSB.

Groups with distinctly different isotopic signatures has important environmental applications. Thousands of heavy-oil wells throughout the WCSB (and probably other basins worldwide) are plagued by unwanted hydrocarbon-gas migration around the well-bores, leading to contamination of soil and shallow aquifers<sup>20,21</sup>. For the first time, the source depth of these gases in the WCSB can be accurately determined using isotopic fingerprints generated through routine analytical procedures (Fig. 4). The use of isotopic fingerprints to determine gas-source depths in our study area has greatly reduced well-remediation costs, and may replace nearly all other logging techniques traditionally used for source detection of leaking gas.

Geochemical analyses continue to be important in elucidating the origin of natural-gas deposits. We have identified two overlying, yet distinct, gas accumulations in the WCSB, and are able to distinguish between the different types of  $\text{C}_{2+}$  gases by using  $\delta^{13}\text{C}$  values. The discovery of thermogenic gas derived at temperatures below 62 °C lowers the limit at which thermal gas accumulation may be expected. Kinetic models based on high-temperature scenarios, however, cannot account for the gas accumulations that have occurred in the Colorado-Group shales. Thus we must look for alternative mechanisms (equilibrium or catalytic) that control low-temperature gas generation. Finally, we have shown that the evolution of gas in two geologically similar environments (the Michigan basin and WCSB) can be very different, and we hope that our findings will lead to geochemical studies of this type (and their applications) in other non-conventional gas deposits worldwide. □

Received 21 September; accepted 15 December 1998.

1. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence: A New Approach to Oil and Gas Exploration* (Springer, Berlin, 1978).
2. Price, L. C. Primary petroleum migration from shales with oxygen-rich organic matter. *J. Petrol. Geol.* **12**, 289–324 (1989).
3. Rich, K., Muehlenbachs, K., Uhrich, K. & Greenwood, G. Carbon isotope characterization of migrating gas in the heavy oil fields of Alberta, Canada. *Soc. Petrol. Eng.* **30265**, 215–224 (1995).
4. Deroo, G., Powell, T. G., Tissot, B. & McCrossan, R. G. The origin and migration of petroleum in the Western Canadian Sedimentary basin, Alberta. *Geol. Surv. Can. Bull.* **262** (1977).
5. Leckie, D. A., Bhattacharya, J. P., Block, J., Gilboy, C. F. & Norris, B. in *Geological Atlas of the Western Canadian Sedimentary Basin* (eds Mossop, G. D. & Shetsen, I.) 335–352 (Can. Soc. Petrol. Geol. Alberta Res. Council, Calgary, 1994).
6. Bustin, R. M. Organic maturity in the western Canada sedimentary basin. *Int. Coal Geol.* **19**, 319–358 (1991).
7. Osadetz, K. G., Pearson, D. E. & Stasiuk, L. Paleogeothermal gradients and changes in the geothermal gradient field of the Alberta Plains. *Curr. Res. C. Geol. Surv. Can. Pap.* **90-1D**, 165–178 (1990).
8. England, T. D. J. & Bustin, R. M. Thermal maturation of the western Canada sedimentary basin south Red Deer River: 1) Alberta plains. *Bull. Can. Petrol. Geol.* **34**, 71–90 (1986).
9. Schoell, M. The hydrogen and carbon isotopic composition of methane from natural gases of various origins. *Geochim. Cosmochim. Acta* **44**, 649–661 (1980).
10. James, A. T. Correlation of natural gas by use of carbon isotopic distribution between hydrocarbon components. *Am. Assoc. Petrol. Geol. Bull.* **67**, 1176–1191 (1983).
11. Clayton, C. Carbon isotope fractionation during natural gas generation from kerogen. *Mar. Petrol. Geol.* **8**, 232–240 (1990).
12. Chung, H. M., Gormly, J. R. & Squires, R. M. Origin of gaseous hydrocarbons in subsurface environments: theoretical considerations of carbon isotope distribution. *Chem. Geol.* **71**, 97–103 (1988).
13. James, A. T. Correlation of reservoir gases using the carbon isotopic compositions of wet gas components. *Am. Assoc. Petrol. Geol. Bull.* **74**, 1441–1458 (1990).
14. Martini, A., Budai, A., Walters, L. & Schoell, M. Economic accumulations of biogenic methane. *Nature* **383**, 153–158 (1996).
15. Fux, A. N. The use of stable carbon isotopes in hydrocarbon exploration. *J. Geochem. Explor.* **7**, 155–188 (1977).
16. Rooney, M. A., Claypool, G. E. & Chung, H. M. Modeling thermogenic gas generation using carbon isotope ratios of natural gas hydrocarbons. *Chem. Geol.* **126**, 219–232 (1995).
17. Mango, F. D., Hightower, J. W. & James, A. T. Role of transition-metal catalysis in the formation of natural gas. *Nature* **368**, 536–538 (1994).
18. Berner, U. & Faber, E. Empirical carbon isotope/maturity relationships for gases from algal kerogens and terrigenous organic matter, based on dry, open-system pyrolysis. *Org. Geochem.* **24**, 947–955 (1996).
19. Butala, S. J. et al. Mineral-catalyzed formation of natural gas during coal maturation. *Am. Assoc. Petrol. Geol. Ann. Conv. Abstr.* **1**, 102 (1998).
20. Erno, B. & Schmitz, R. in *Tech. Meeting Petrol. Soc. Can. Inst. Mining Vol. 94* paper 73 (1994).
21. Taylor, S. W., Sherwood-Lollar, B., Wasenaar, L. & Hendry, M. J. Bacterial hydrocarbon production in a shallow bedrock aquifer and surficial till aquitards, Saskatchewan, Canada. *Geol. Soc. Am. Abstr.* **29**, 186 (1997).

**Acknowledgements.** We thank K. Muehlenbachs, S. Talman and A. T. James for discussions that helped to improve the original manuscript; L. C. Price for reviews of the first draft; and Husky Oil Operation Ltd for providing samples. The Canadian Association of Petroleum Producers supported this research.

Correspondence and requests for materials should be addressed to Karlis Muehlenbachs (e-mail: kmuehlen@gpu.srv.ualberta.ca).

# Evidence for motion between Nubia and Somalia along the Southwest Indian ridge

Dezhi Chu\*† & Richard G. Gordon\*

\* Department of Geology & Geophysics, Rice University, Houston, Texas 77005, USA

The East African rift marks the northern boundary of the Nubian (West African) and Somalian (East African) plates, and has formed by horizontal stretching due to the separation of these plates<sup>1</sup>. South of ~20° S, any expression of deformation or seismicity due to the relative motion of these two distinct plates vanishes, although the boundary must continue until it intersects another plate boundary. The nearest such boundary is that of the Antarctic plate, marked by the Southwest Indian ridge. But previous analyses of plate-motion data have indicated no significant difference between Nubia–Antarctica and Somalia–Antarctica motion<sup>2,3</sup>. Here we show, using a large compilation of plate-motion data, that Nubia–Antarctica motion does differ from Somalia–Antarctica motion, and we determine a relative angular velocity of the two plates that has compact confidence limits. Our analysis places the pole of rotation near to the southern limit of African seismicity, implying that the southern part of the Nubian–Somalian plate boundary is a diffuse zone of convergence (up to ~2 mm yr<sup>-1</sup>), whereas up to ~6 mm yr<sup>-1</sup> of separation is accommodated across the East African rift—about half the separation rate of the slowest mid-ocean ridge.

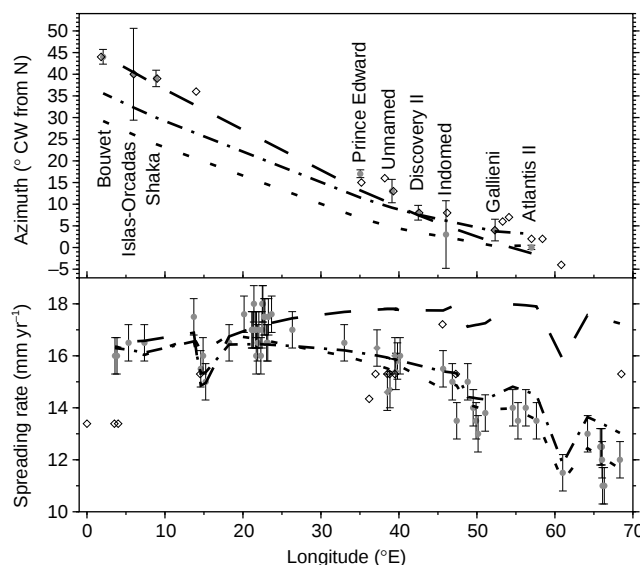
Missing from previous investigations of Nubia–Somalia motion has been the resolution of any significant difference between Nubia–Antarctica and Somalia–Antarctica motion. Here we use a new set of 59 rates of sea-floor spreading since the middle of chron 2A (3.2 Myr ago), many more than the 12 rates used before<sup>2,4</sup>. Marine magnetic profiles obtained from the US National Geophysical Data Center, which were unavailable to previous studies<sup>2,4</sup>, indicate that rates are nearly 3 mm yr<sup>-1</sup> faster than previously estimated along the southwestern part of the Southwest Indian ridge (SWIR) near the Bouvet triple junction, and ~3 mm yr<sup>-1</sup> slower than previously estimated along the northeastern part of the SWIR near the Rodrigues triple junction (Fig. 1). We use azimuths of transform faults<sup>5</sup> that are similar to those used before<sup>1</sup>, the main exception being the important addition of a well constrained azimuth for the Atlantis II transform fault from a SeaBeam (multibeam sonar) survey<sup>6</sup> (Fig. 1).

We apply an *F*-ratio test to determine whether the plate motion data along the SWIR are fitted significantly better by a model with separate Nubian and Somalian plates than by a model with a single African plate<sup>7,8</sup>. In this test we assume that the boundary between Nubia and Somalia is narrow where it intersects the SWIR, that is, that the intersection lies between the locations of adjacent plate-motion data. First, we consider a data set consisting only of spreading rates along the SWIR; the improvement in fit to the data is only marginally statistically significant (0.051 significance level or, equivalently, the 94.9% confidence level). Second, we consider a data set consisting only of azimuths of transform faults along the SWIR; the improvement in fit is statistically significant (0.029 significance level). Third, we consider a data set consisting of both transform-fault azimuths and spreading rates along the SWIR; the improvement in fit is highly significant ( $2 \times 10^{-16}$  significance level). The best location for the intersection of a hypothetical narrow Nubia–Somalia boundary with the SWIR is between the

Prince Edward transform fault (35.05° E), the azimuth of which was determined from many bathymetric profiles, and the intersection of the SWIR and a magnetic anomaly profile (37.2° E), from which a spreading rate was determined. The 95% confidence interval is from 35.05° E to 38.50° E along the SWIR.

Although we cannot conclusively demonstrate that the boundary between Nubia and Somalia is not narrow where it intersects the SWIR, we believe it is wide and diffuse for several reasons. (1) There is no expression in the topography (Fig. 2) or satellite-derived gravity of a narrow boundary that intersects the SWIR within the 95% confidence interval for a hypothetical narrow plate boundary (that is, between 35.05° E and 38.50° E) or elsewhere. (2) There is no narrow band of earthquakes in the sea floor southeast of southern Africa, and certainly no band that intersects the SWIR near the confidence interval for a hypothetical narrow boundary (Fig. 2). (3) The only recorded earthquake having a magnitude exceeding 5.5 is a thrust event (the 26 February 1988 event, moment magnitude  $M_w = 6.8$ ) located near the SWIR at 37.30° S, 47.97° E, far to the northeast of the confidence interval for a hypothetical narrow boundary. (4) The Nubia–Somalia angular velocity calculated from data along the SWIR assuming that the boundary is narrow differs significantly (0.001 significance level) from that estimated from plate motion data in the Red Sea and the Gulf of Aden, whereas that assuming a diffuse boundary, as discussed below, does not (16% significance level).

The plate-motion data suggest possible limits for a diffuse plate



**Figure 1** Comparisons of observed rates of sea-floor spreading and azimuths with those calculated from various models. Dash-dotted curves, rates and azimuths calculated from the angular velocity that best fits all the data along the SWIR (calculated as if there were a single African plate and not distinct Nubian and Somalian plates). Long-dashed curves, rates and azimuths calculated from the angular velocity (1.9° N, 38.5° W; 0.162° Myr<sup>-1</sup>) that best fits the data west of 32.99° E along the SWIR, corresponding to motion between the Nubian and Antarctic plates. Short-dashed curves, rates and azimuths calculated from the angular velocity (19.5° N, 65.6° W; 0.154° Myr<sup>-1</sup>) that best fits the data east of 52.32° E along the SWIR, corresponding to motion between the Somalian and Antarctic plates. Azimuths we use here are shown by filled circles. The new set of 59 rates of sea-floor spreading (available from the authors on request) is shown by filled symbols; circles if determined from digital data, diamonds if from analogue data. Open diamonds show the rates and azimuths used for plate-motion model NUVEL-1A<sup>4,23</sup>. Vertical bars show 1σ uncertainties. Rates were assigned an uncertainty of 0.7 mm yr<sup>-1</sup> based on their standard deviation about best-fitting models. Azimuths of transform faults and their uncertainties are from ref. 5; all azimuths were determined from bathymetric data. We use no transform azimuths estimated from altimetric data because of their inconsistencies<sup>5</sup>.

† Present address: Exxon Production Research Company, PO Box 2189, Houston, Texas 77252, USA.

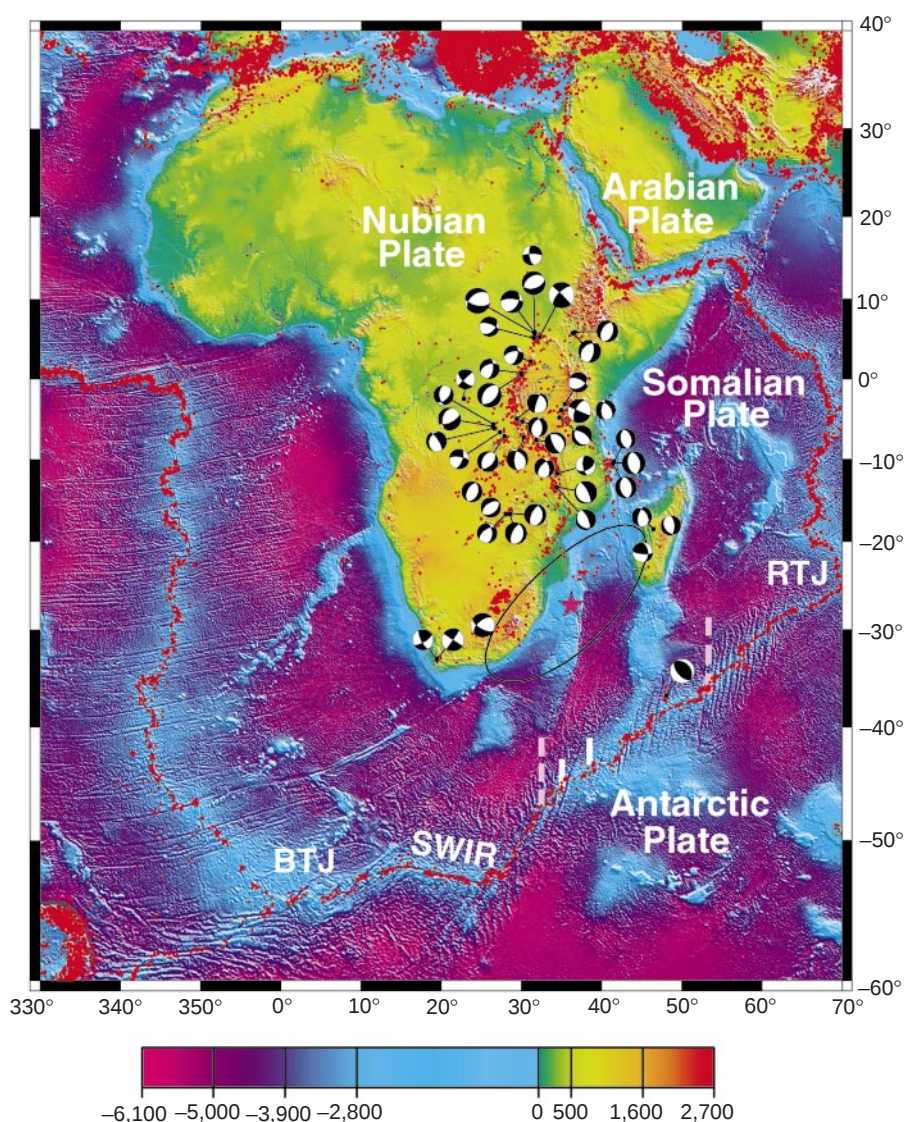


boundary. Spreading rates observed east of, and including,  $32.99^{\circ}$  E are systematically slower than predicted by a Nubian–Antarctic plate motion model fit to data west of  $32.99^{\circ}$  E; also, transform azimuths observed west of, and including, the Gallieni transform fault ( $52.3^{\circ}$  E) are systematically clockwise (CW) of those predicted by a Somalian–Antarctic plate motion model fit to data east of  $52.3^{\circ}$  E (Fig. 1). These misfits could be explained by the existence of a diffuse plate boundary that intersects the SWIR from  $32.99^{\circ}$  E to  $52.32^{\circ}$  E (Fig. 2). We estimate the angular velocity of Nubia relative to Antarctica from the data west of  $32.99^{\circ}$  E along the SWIR; and we estimate the angular velocity of Somalia relative to Antarctica from the data east of  $52.32^{\circ}$  E along the SWIR. Taking the difference between Nubia–Antarctica and Arabia–Antarctica angular velocities gives an estimate of the angular velocity of Nubia relative to Somalia from data only along the SWIR (Fig. 3).

Motion between Nubia and Somalia can be also estimated by taking the difference between their motion relative to Arabia. For

the angular velocity of Nubia relative to Arabia, we use the value recently determined from 45 spreading rates from the Red Sea<sup>9</sup>. For the angular velocity between Arabia and Somalia, we estimated 46 spreading rates and the azimuths of 4 transform faults and obtained an angular velocity similar to previous estimates<sup>4,10</sup>, but with smaller confidence limits. The difference between the Nubia–Arabia and Arabia–Somalia angular velocities gives an estimate of the angular velocity of Nubia relative to Somalia that is independent of data along the SWIR (Fig. 3).

When the SWIR and circum-Arabia plate-motion data sets (again omitting data from  $32.99^{\circ}$  E to  $52.32^{\circ}$  E along the SWIR) are combined, a pole with a compact confidence region is found off the east coast of South Africa,  $\sim 10^{\circ}$  south of the southern limit of large normal-faulting earthquakes in Africa, and near the transition between continental and oceanic crust (Figs 2 and 3). This pole of rotation, which is used in the rest of this Letter, differs distinctly and significantly from previous poles<sup>3,11</sup> (Fig. 3), which indicated that



**Figure 2** Geometry, topography, seismicity, earthquake mechanisms, and pole of rotation for the Nubian and Somalian plates. Colour scale gives elevation in metres above sea level. Small red circles, earthquake locations obtained from the US National Earthquake Information Center. Mechanisms are shown for earthquakes with magnitudes  $>5.5$  and are from refs 24–27 and Harvard CMT solutions. Abbreviations: BTJ, Bouvet triple junction; RTJ, Rodrigues triple junction;

tion; SWIR, Southwest Indian ridge. North–south white line segments adjacent to the SWIR show the eastern and western confidence limits on the intersection of a hypothetical narrow Nubia–Somalia boundary with the SWIR. North–south pink line segments show the limits we assumed for our proposed diffuse Nubian–Somalian plate boundary along the SWIR.

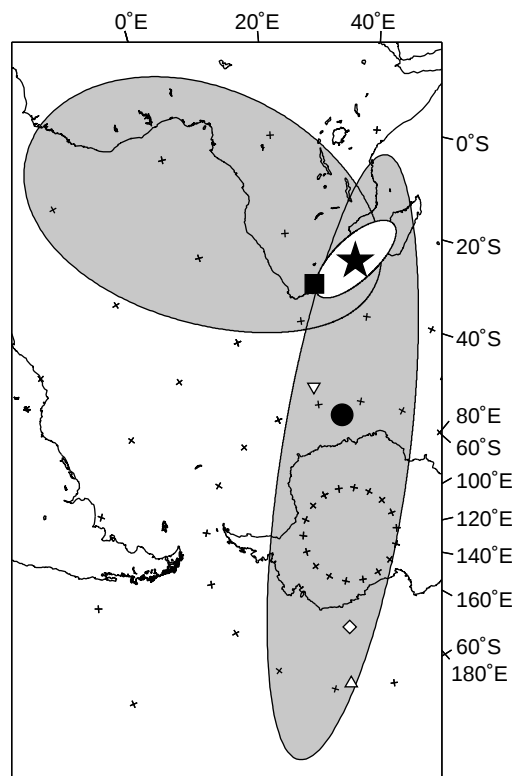


stretching between Nubia and Somalia continues southeastwards to the SWIR. The southern limit of large normal-faulting earthquakes and the East Africa rift is 1,000–1,500 km north of this newly determined pole of rotation, near which stretching must become vanishingly small (Fig. 2). Moreover, the location of this pole unexpectedly predicts that southwest–northeast convergence is accommodated in the oceanic lithosphere between Africa and the SWIR.

The thrust event near the SWIR shows that the observed sense of deformation is southwest–northeast shortening, consistent with the predicted convergence (Fig. 2). The new Nubia–Somalia angular velocity predicts that motion across a hypothetical narrow Nubia–Somalia boundary that includes the location of the earthquake would be  $2.4 \pm 0.6 \text{ mm yr}^{-1}$  towards  $\text{N}42^\circ\text{E} \pm 38^\circ$ . (All uncertainties following ‘ $\pm$ ’ signs are 95% confidence limits.) The angular velocity also predicts motion of Nubia relative to Somalia across the northern East Africa rift (at  $10^\circ\text{N}$ ,  $40^\circ\text{E}$ ) of  $6.0 \pm 1.5 \text{ mm yr}^{-1}$  towards  $\text{N}84^\circ\text{W} \pm 9^\circ$ .

Figure 2 shows that many moderate to large earthquakes occur in the continental lithosphere north of the pole of rotation, whereas few occur in the oceanic lithosphere southeast of this pole. This seems to be mainly a consequence of this pole being a greater distance from much of the East Africa rift than it is from any part of the zone of convergence. There is a substantial region both north and southeast of this pole that lacks large earthquakes or other obvious indicators of active deformation.

The horizontal contraction in the oceanic lithosphere southeast



**Figure 3** Nubia–Somalia rotation poles and confidence regions. Black symbols show rotation poles determined here: circle, from Red Sea and Gulf of Aden plate-motion data (with 95% confidence ellipsoid); square, from SWIR plate-motion data (with 95% confidence ellipsoid; data from  $32.99^\circ\text{E}$  to  $52.32^\circ\text{E}$  have been omitted); star, from these two data sets combined (with 95% confidence ellipsoid). The angular velocity from the combined data set is  $0.089^\circ\text{Myr}^{-1}$  about  $27.3^\circ\text{S}$ ,  $36.2^\circ\text{E}$ . The  $xx$ ,  $xy$ ,  $xz$ ,  $yy$ ,  $yz$  and  $zz$  elements of its covariance matrix are respectively 70, 18, –61, 136, 26 and 123 in units of  $10^{-10} \text{ rad}^2 \text{ Myr}^{-2}$ . White symbols show other Nubia–Somalia rotation poles: diamond, from ref. 11; triangle, from ref. 3 (Nubia–Arabia–Somalia circuit); inverted triangle, from ref. 3 (four-plate Nubia–Arabia–Somalia–Antarctica system).

of Africa is manifested differently from that elsewhere in the Indian Ocean. In the central Indian and Wharton basins, where topography and gravity indicate the presence of  $\sim 200\text{-km}$ -wavelength undulations with peak-to-trough amplitudes of  $\sim 1 \text{ km}$ , the entire lithosphere is buckling in response to compressional stresses<sup>12–18</sup>. In contrast, there is little evidence for lithospheric buckling between Africa and the SWIR. Martinod and Molnar<sup>14</sup> have calculated that a force per unit length of  $(4.8 \pm 1.3) \times 10^{12} \text{ N m}^{-1}$  could fold the lithosphere where sediments fill the troughs in the folds, and that a force per unit length of  $(1.1 \pm 0.3) \times 10^{13} \text{ N m}^{-1}$  could do this where sediments are absent. Therefore the minimum force per unit length is  $\sim 4.8 \times 10^{12} \text{ N m}^{-1}$  in the Central Indian basin where sediments fill the troughs, and is  $\sim 1.1 \times 10^{13} \text{ N m}^{-1}$  in the Wharton basin where sediments are sparse. Similarly, it follows that the force per unit length driving the contraction of sea floor southeast of Africa, where sediment cover is also sparse, is less than  $\sim 1.1 \times 10^{13} \text{ N m}^{-1}$ . The driving force is surely greater in the Wharton basin and probably greater in the Central India basin because of the presence of nearby subducting slabs attached to the Indo–Australian composite plate, as well as the outward push of the Tibetan plateau. The more modest forces driving the deformation southeast of Africa are probably due to the outward push from the elevated regions of Afar and the East Africa rift.

The Nubian–Somalian diffuse plate boundary, like the Indian–Capricorn<sup>18,19</sup>, Capricorn–Australian<sup>18</sup>, and North American–South American<sup>4,20</sup> boundaries, consists of two distinct portions, one accommodating convergence and one accommodating divergence. The pole of rotation lies between these two distinct portions (Fig. 2), which indicates that the two plates are strongly coupled dynamically across their boundary<sup>20</sup>. Thus, the African composite plate is in an important sense a single mechanical entity, although Nubian and Somalian component plates are in resolvable relative motion.

The non-closure of the plate circuit about the Rodrigues triple junction, that is, through the Indo–Australian, Antarctic and African plates, was demonstrated by Minster and Jordan<sup>21</sup>. At that time, it seemed reasonable to expect a simple explanation for the non-closure, such as a single diffuse plate boundary that intersected one of the three plate boundaries that meet at the triple junction: the Central Indian ridge, the Southeast Indian ridge and the SWIR. With the recognition here of a probably diffuse plate boundary that intersects the SWIR, the recognition last year of a diffuse plate boundary that intersects the Southeast Indian ridge<sup>18</sup>, and the long-standing recognition of one that intersects the Central Indian ridge<sup>22</sup>, it has now been shown that a plate boundary intersects each of the three narrow plate boundaries that meet at the Rodrigues triple junction. Simplicity has not been a good guide in predicting the tectonics of the Indian Ocean. □

Received 19 June; accepted 11 December 1998.

- McKenzie, D. P., Davies, D. & Molnar, P. Plate tectonics of the Red Sea and East Africa. *Nature* **226**, 243–248 (1970).
- DeMets, C., Gordon, R. G. & Argus, D. F. Intraplate deformation and closure of the Australia–Antarctica–Africa plate circuit. *J. Geophys. Res.* **93**, 11877–11897 (1988).
- Jestin, F., Huchon, P. & Gaulier, J. M. The Somalia plate and the East African Rift System: present-day kinematics. *Geophys. J. Int.* **116**, 637–654 (1994).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
- Spitzak, S. & DeMets, C. Constraints on present-day plate motions south of  $30^\circ\text{S}$  from satellite altimetry. *Tectonophysics* **253**, 167–208 (1996).
- Dick, H. J. B. *et al.* Tectonic evolution of the Atlantis II Fracture Zone. *Proc. ODP Sci. Res.* **118**, 359–398 (1991).
- Stein, S. & Gordon, R. G. Statistical tests of additional plate boundaries from plate motion inversions. *Earth Planet. Sci. Lett.* **69**, 401–412 (1984).
- Gordon, R. G., Stein, S., DeMets, C. & Argus, D. F. Statistical tests for closure of plate motion circuits. *Geophys. Res. Lett.* **14**, 587–590 (1987).
- Chu, D. & Gordon, R. G. Current plate motions across the Red Sea. *Geophys. J. Int.* **135**, 313–328 (1998).
- Gordon, R. G. & DeMets, C. Present-day motion along the Owen Fracture Zone and Dalrymple Trough in the Arabian Sea. *J. Geophys. Res.* **94**, 5560–5570 (1989).
- Chase, C. G. Plate kinematics: The Americas, East Africa, and the rest of the world. *Earth Planet. Sci. Lett.* **37**, 355–368 (1978).
- Weissel, J. K., Anderson, R. N. & Geller, C. A. Deformation of the Indo–Australian plate. *Nature* **287**, 284–291 (1980).

13. McAdoo, D. C. & Sandwell, D. T. Folding of oceanic lithosphere. *J. Geophys. Res.* **90**, 8563–8569 (1985).
14. Martinod, J. & Molnar, P. Lithospheric folding in the Indian Ocean and the rheology of the oceanic plate. *Bull. Soc. Géol. Fr.* **166**, 813–821 (1995).
15. Zuber, M. T. & Parmentier, E. M. Finite amplitude folding of a continuously viscosity-stratified lithosphere. *J. Geophys. Res.* **101**, 5489–5498 (1996).
16. Petroy, D. E. & Wiens, D. A. Historical seismicity and implications for diffuse plate convergence in the northeast Indian Ocean. *J. Geophys. Res.* **94**, 12301–12319 (1989).
17. Stein, C. A., Cloetingh, S. & Wortel, R. Seasat-derived gravity constraints on stress and deformation in the northeastern Indian Ocean. *Geophys. Res. Lett.* **16**, 823–826 (1989).
18. Royer, J.-Y. & Gordon, R. G. The motion and boundary between the Capricorn and Australian plates. *Science* **277**, 1268–1274 (1997).
19. Gordon, R. G., DeMets, C. & Argus, D. F. Kinematic constraints on distributed lithospheric deformation in the equatorial Indian Ocean from present motion between the Australian and Indian plates. *Tectonics* **9**, 409–422 (1990).
20. Gordon, R. G. The plate tectonic approximation: Plate nonrigidity, diffuse plate boundaries, and global plate reconstructions. *Annu. Rev. Earth Planet. Sci.* **26**, 615–642 (1998).
21. Minster, J. B. & Jordan, T. H. Present-day plate motions. *J. Geophys. Res.* **83**, 5331–5354 (1978).
22. Wiens, D. A. *et al.* A diffuse plate boundary model for Indian Ocean tectonics. *Geophys. Res. Lett.* **12**, 429–432 (1985).
23. DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motion. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
24. Sykes, L. R. Mechanism of earthquakes and nature of faulting on the mid-oceanic ridges. *J. Geophys. Res.* **72**, 2131–2153 (1967).
25. Shudofsky, G. N. Source mechanisms and focal depths of east African earthquakes using Rayleigh-wave inversion and body-wave modelling. *Geophys. J. R. Astron. Soc.* **83**, 563–614 (1985).
26. Grimson, N. L. & Chen, W.-P. Earthquakes in the Davie Ridge-Madagascar region and the southern Nubian-Somalian plate boundary. *J. Geophys. Res.* **93**, 10439–10450 (1988).
27. Wagner, G. S. & Langston, C. A. East Africa earthquake body wave inversion with implications for continental structure and deformation. *Geophys. J.* **94**, 503–518 (1988).

**Acknowledgements.** We thank F. Pollitz, T. Henstock and A. Gripp for comments and advice and A. Gripp for assistance in compiling earthquake mechanisms. This work was supported by NASA and the NSF.

Correspondence and requests for materials should be addressed to R.G.G. (e-mail: rgg@rice.edu).

## Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation

Paul A. Garriss\*, Michaux Kilpatrick†, Melissa A. Bunin†, Darren Michael†, Q. David Walker† & R. Mark Wightman†

\* Department of Biological Sciences, 102 Felmley Hall, Illinois State University, Normal, Illinois 61790-4120, USA

† Department of Chemistry and Curriculum in Neurobiology, Venable Hall, CB 3290, University of North Carolina, Chapel Hill, North Carolina 27599-3290, USA

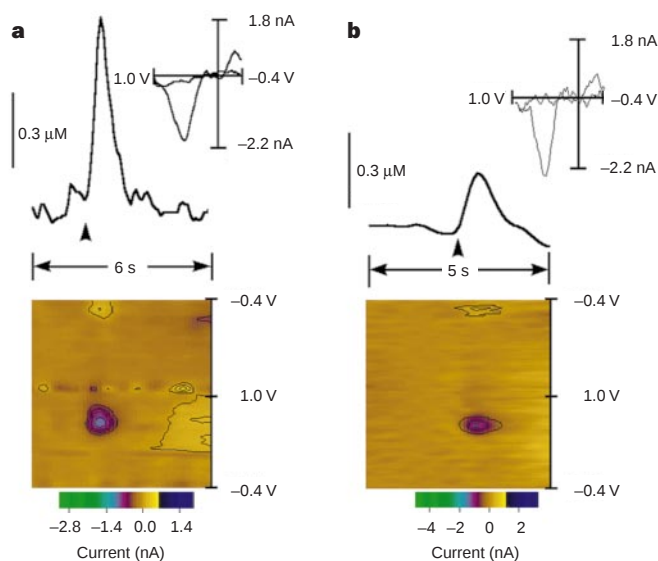
Mesolimbic dopamine-releasing neurons appear to be important in the brain reward system<sup>1,2</sup>. One behavioural paradigm that supports this hypothesis is intracranial self-stimulation (ICS), during which animals repeatedly press a lever to stimulate their own dopamine-releasing neurons electrically<sup>3–6</sup>. Here we study dopamine release from dopamine terminals in the nucleus accumbens core and shell in the brain by using rapid-responding voltammetric microsensors<sup>7</sup> during electrical stimulation of dopamine cell bodies in the ventral tegmental area/substantia nigra brain regions. In rats in which stimulating electrode placement failed to elicit dopamine release in the nucleus accumbens, ICS behaviour was not learned. In contrast, ICS was acquired when stimulus trains evoked extracellular dopamine in either the core or the shell of the nucleus accumbens. In animals that could learn ICS, experimenter-delivered stimulation always elicited dopamine release. In contrast, extracellular dopamine was rarely observed during ICS itself. Thus, although activation of mesolimbic dopamine-releasing neurons seems to be a necessary condition for ICS, evoked dopamine release is actually diminished during ICS. Dopamine may therefore be a neural substrate for novelty<sup>8</sup> or reward expectation<sup>9</sup> rather than reward itself.

Dopamine neurotransmission is thought to be involved in ICS because, in addition to the anatomical coincidence of ICS sites and

mesolimbic dopamine neurons, injection of dopaminergic agents into the nucleus accumbens, the major projection field of dopamine neurons originating in the ventral tegmental area (VTA), modifies ICS<sup>10</sup>. Moreover, selective lesions of dopaminergic neurons decrease ICS<sup>11</sup>. Despite this, it is unclear whether the increase in extracellular dopamine measured in the nucleus accumbens during ICS sessions<sup>12–15</sup> arises from direct activation of mesolimbic dopamine neurons or from indirect activation<sup>16</sup>, because, until recently<sup>17</sup>, transient chemical changes during behaviour could not be measured. To resolve this controversy, we monitored extracellular dopamine in the nucleus accumbens during transient ICS of the VTA/substantia nigra, using stimulating electrodes positioned to depolarize dopamine neurons directly. The cyclic voltammetric results, displayed in colour<sup>18</sup>, allow all of the detected species, including dopamine, to be observed simultaneously. Thus, the errors in interpretation resulting from poor chemical specificity and sensitivity that affected some previous uses of *in vivo* electrochemistry can be avoided<sup>19,20</sup>.

Male Sprague–Dawley rats were prepared for ICS by implanting stimulating electrodes unilaterally in the VTA/substantia nigra. During surgery, a train of pulses with characteristics similar to those supporting ICS was delivered to the electrode while extracellular dopamine in the nucleus accumbens shell or core was monitored using a carbon-fibre microelectrode<sup>17</sup>. Whether or not dopamine was observed, the stimulating electrode was cemented in place near the stereotaxic coordinates of the VTA/substantia nigra.

After a few days for recovery and ICS training, a fresh microelectrode was lowered into the nucleus accumbens, and evoked dopamine release was studied in freely moving animals. In animals that exhibited ICS, the operator-delivered stimulus elicited dopamine release (examples from the nucleus accumbens core and shell are shown in Fig. 1a, b). The colour representation of the voltammetric current shows the emergence of the oxidation peak for dopamine at



**Figure 1** Dopamine release in the nucleus accumbens evoked by a single, experimenter-delivered stimulus train (0.4 s, 60 Hz). **a**, Release in nucleus accumbens core; stimulus conditions: pulse width 2 ms, amplitude 188  $\mu$ A. **b**, Release in nucleus accumbens shell; stimulus conditions: pulse width 1 ms, amplitude 125  $\mu$ A. Bottom, topographic representation of the voltammetric data, with current changes encoded in colour. The currents for the first 15 scans were subtracted from the remainder to reveal changes. Top, dopamine concentration change determined from electrode calibration after *in vivo* use and corrected for the pH change<sup>18</sup>. The arrowheads indicate stimulus delivery. Inset, the cyclic voltammogram obtained at the end of the stimulus is identical to the cyclic voltammogram for dopamine.

0.6 V, coupled to the reduction of electroformed dopamine-o-quinone at  $-0.1$  V, during the stimulation interval. Broader features were often seen at later times on both portions of the voltammetric sweeps (Fig. 1a). These features arise from pH or other ionic changes that follow electrical stimulation. These contributions can be subtracted<sup>18</sup>, and the remaining signal reveals the dopamine change with time. The background-subtracted cyclic voltammograms obtained at the end of the stimulations (Fig. 1) further identify the presence of dopamine. The different time courses of the dopamine signals in each region are a direct measure of the differences in release and uptake<sup>21</sup>.

Transient stimulation trains of the type used here for ICS directly depolarize dopamine neurons, consequently evoking the release of dopamine. Although single trains produced no obvious behavioural responses, the transient concentrations of extracellular dopamine ( $\sim 150$  nM) observed were quite high compared with the levels detected with microdialysis probes<sup>22</sup>. The small size of the carbon-fibre microelectrode allows positioning of the probe directly in the midst of release sites, enabling detection of the functionally relevant concentration of dopamine. Thus, dopamine nerve terminals in each region can respond to reinforcing stimuli that a rat would self-administer by releasing dopamine.

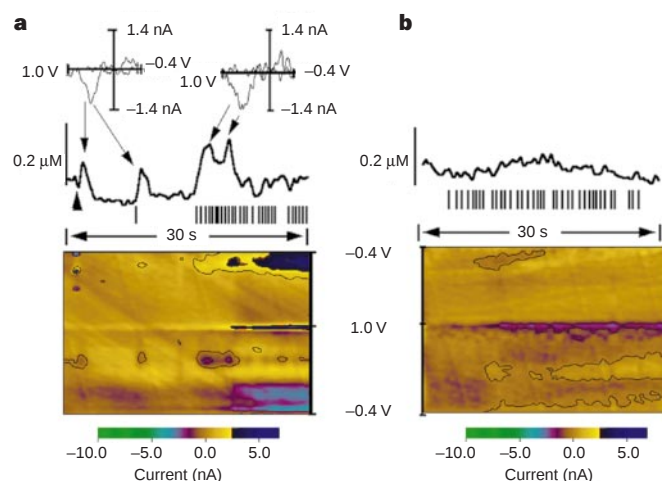
Of the animals in which robust dopamine release was observed in either region of the nucleus accumbens, all successfully learned to press a bar to undergo ICS ( $n = 6$  for each region). In some animals ( $n = 4$ ) in which dopamine release was not evoked, ICS could not be achieved despite identical training. These four animals did not show the exploratory behaviour exhibited by the animals in which the stimulus trains evoked dopamine release. These results support the concept that evoked dopamine release is associated with the acquisition of ICS.

Following experimenter-applied stimulation, dopamine amounts were monitored during ICS. Dopamine was rarely observed (in 3 out of 36 trials in the nucleus accumbens core and 3 out of 12 trials in the nucleus accumbens shell), even though each bar press caused a stimulus train with parameters that were identical to those

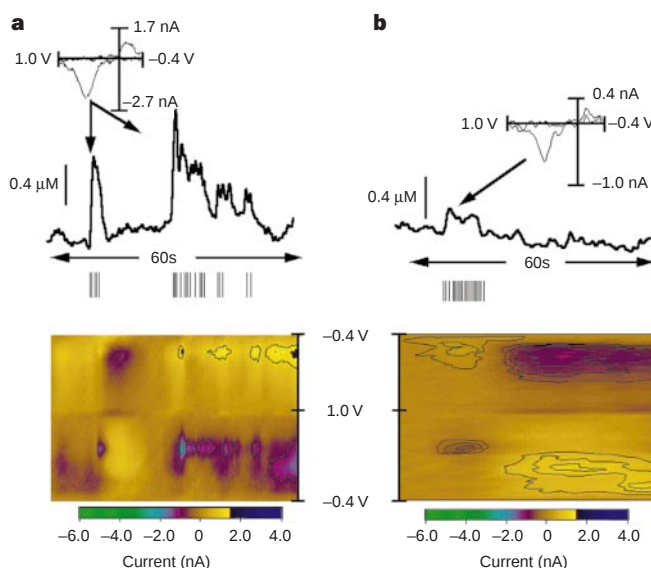
delivered by the experimenter immediately before ICS when dopamine was observed. Examples from measurements taken in the nucleus accumbens shell of an animal in which dopamine was observed are shown in Fig. 2. In Fig. 2a, the 100 nM dopamine signal evoked by a single, operator-delivered bar press is apparent, as is the dopamine release evoked by the first bar press and during the initial ICS period. However, as ICS continued, extracellular dopamine declined to unmeasurable levels. In a second ICS trial 30 min later, bar-press rates were vigorous but extracellular dopamine was not observed. Although the colour representation shows that changes in the voltammetric signal occurred, these changes were not at the potentials of dopamine oxidation and re-reduction. This trial is representative of most of the trials.

In contrast to the results obtained during ICS, marked dopamine release was evoked in untrained animals during playback of a previous animal's bar-press record in every case tested ( $n = 6$ ). Similar responses were observed with repetitive trains delivered by the experimenter. In the first trial, dopamine increased to a mean level of  $1.8 \pm 1.06$   $\mu$ M, a level more than 13 times greater than observed in this region during ICS ( $P \leq 0.05$ ). However, as shown in the example trace in Fig. 3, the initial dopamine rise was not maintained and decreased, although the electrical stimulations were continually delivered. In all cases, dopamine release occurred in the second trial but was attenuated as in the example in Fig. 3. During ICS, dopamine release was observed in a second trial in only one animal.

Because dopamine neurotransmission accompanying impulse flow occurs through diffusion in the extracellular space in the nucleus accumbens, these measurements provide a functional index of dopamine neuronal signalling in the brain<sup>23,24</sup>. The initial stimulus applied to an untrained animal, whether alone or as part of another animal's bar-press record, always evoked dopamine release in the nucleus accumbens. In contrast, dopamine release was observed only infrequently during ICS. With similar stimulus parameters, the absence of ICS-evoked dopamine release has also been observed when using *in vivo* microdialysis<sup>13</sup>. When dopamine



**Figure 2** Voltammetric changes recorded in the nucleus accumbens shell in a single animal during ICS. **a**, Initial ICS trial. **b**, ICS trial in the same location 30 min later. Top, dopamine concentration change and cyclic voltammograms. An operator-delivered stimulus train is indicated by the arrowhead in **a**. Each vertical bar indicates a lever press. Bottom, topographical data. A non-dopaminergic change is seen in the last quarter of the interval shown in **a**, and during much of **b**.



**Figure 3** Voltammetric changes measured in the nucleus accumbens shell of an untrained animal during the playback of another animal's ICS record. **a**, Initial recording. **b**, Recording 30 min later in the same animal. Vertical bars indicate lever presses. Top, dopamine concentration change, corrected for the pH change and calibrated as in Fig. 1. Bottom, contour plots reveal an increase in dopamine during the time interval in which the animal is being stimulated, as well as a pH change. Insets, the voltammograms confirm that dopamine release is being monitored.



was seen during ICS, release occurred early in the set of ICS trials; it was attenuated relative to that evoked by non-contingent stimulation, but, as with non-contingent stimulation, it diminished during continuous stimulations. This failure to observe consistent dopamine release while directly depolarizing dopamine neurons shows that the brain has potent mechanisms that suppress release from mesolimbic dopamine neurons. These mechanisms may originate from inhibitory actions of dopamine at its terminal autoreceptors or other transmitter-receptor interactions. Prolonged, intense stimulation in anaesthetized animals<sup>25</sup> can deplete releasable dopamine stores, and this may also be important here. Irrespective of mechanism, these results indicate that there is no correlation between continual bar pressing during ICS and increased dopaminergic neurotransmission in the nucleus accumbens.

Although previous work has established that dopamine neurotransmission is a component of brain reward, our results are consistent with evidence that the dopaminergic component is not associated with the hedonistic or 'pleasure' aspects of reward<sup>26</sup>. For example, if a reward in the form of an appetitive liquid is delivered to a monkey in an unpredictable way, dopamine neurons in the substantia nigra increase their rate of firing<sup>27</sup>. However, if the animal learns that a tone precedes the delivery of liquid, dopamine neurons increase their firing rate in response to the tone, not the reward. Likewise, the rewarding effects of cocaine do not require dopamine; mice lacking the gene for the dopamine transporter, a major target of cocaine, will self-administer cocaine<sup>28</sup>. However, increased dopamine neurotransmission in the nucleus accumbens shell is seen when rats are transiently exposed to a new environment<sup>8</sup>. The increase in extracellular dopamine quickly returns to normal levels and remains there during continued exploration of the new environment. Thus, our results obtained during ICS, an unnatural stimulus, as well as other results obtained during food presentation and exposure to new environments, all natural stimuli, indicate that dopamine release in the nucleus accumbens is related to novelty, predictability or some other aspect of the reward process, rather than to hedonism itself. We speculate that these are the components that are modified by dopamine pharmacological agents with well-documented effects on ICS<sup>4,29,30</sup>. An understanding of the underlying mechanisms that downregulate dopamine release may provide the first opportunity to define the link between transient biochemical regulation of neuronal processes and behaviour. □

## Methods

**Animal use.** Animal care was in accordance with standard procedures and was approved by IACUC of the University of North Carolina. Rats were anaesthetized with either equithesin (0.3 ml per kg body weight intraperitoneally (i.p.)) or a combination of ketamine (85 mg per kg body weight i.p.) and xylazine (5 mg per kg body weight i.p.) and immobilized in a stereotaxic frame. The reference (Ag/AgCl) and auxiliary electrodes for electrochemistry were permanently implanted directly beneath the skull and connected to a pedestal that contained the head-mounted voltammetric amplifier<sup>17</sup>. A bipolar stimulating electrode with tips separated by 1 mm was positioned in the VTA/substantia nigra region (coordinates given in millimetres relative to the bregma with flat skull orientation: -5.6 anteroposterior, 1.0 mediolateral, 8–9 dorso-ventral). Stimulation at this position evokes dopamine release throughout the nucleus accumbens. A guide cannula, permanently mounted above the nucleus accumbens core (1.2 anteroposterior, 1.4 mediolateral) or shell (1.7 anteroposterior, 0.7 mediolateral), held the micromanipulator that was used to position a fresh, cylindrically shaped carbon-fibre microelectrode for each recording session, a procedure that did not require anaesthesia<sup>17</sup>.

Rats trained for ICS were placed in a chamber containing a lever that, when depressed, triggered delivery of a stimulus train to the electrode in the VTA/substantia nigra region. The room lights were turned off and a small bulb in the chamber was illuminated to signal that reinforcement was available. When the animal approached the lever, single trains of pulses were applied by the experimenter. When the animal associated the stimulus with a lever press, ICS was allowed for a 2-min interval. This training was repeated at least twice on

three different days before voltammetric recordings were attempted. Successfully trained animals exhibited bar-press rates of between 1 and 2 per second when stimulus amplitudes of 80–250  $\mu$ A were used. As the delivery of a stimulus train in response to a lever press was delayed by a maximum of 100 ms to avoid simultaneous application of a stimulus pulse and acquisition of a voltammogram and because of the train duration (0.4 s) for ICS, the maximum rate of train application allowed by the hardware was 2 per second. Voltammetric recordings were made during at least two consecutive ICS sessions per animal separated by 30-min intervals.

**Electrochemistry.** Cyclic voltammograms (300 V s<sup>-1</sup> scan rate, potential range -0.4 to 1.0 V versus Ag/AgCl) were collected at 100-ms intervals<sup>17</sup>. Stimulation trains were constant current, 60 Hz, 0.4-s duration and consisted of 1-ms biphasic pulses unless otherwise stated.

Received 26 August; accepted 29 December 1998.

- Schulteis, G., Markou, A., Cole, M. & Koob, G. F. Decreased brain reward produced by ethanol withdrawal. *Proc. Natl Acad. Sci. USA* **92**, 5880–5884 (1995).
- Pich, E. M. et al. Common neural substrates for the addictive properties of nicotine and cocaine. *Science* **275**, 83–86 (1997).
- Shizgal, P. & Murray, B. in *The Neuropharmacological Basis of Reward* (eds Lieberman, J. M. & Cooper, S. J.) 106–162 (Clarendon, Oxford, 1989).
- Wise, R. A. Addictive drugs and brain stimulation reward. *Annu. Rev. Neurosci.* **19**, 319–340 (1996).
- Phillips, A. G. & Fibiger, H. C. in *The Neuropharmacological Basis of Reward* (eds Lieberman, J. M. & Cooper, S. J.) 66–104 (Clarendon, Oxford, 1989).
- Olds, J. & Milner, P. M. Positive reinforcement produced by electrical stimulation of septal area and other regions of the rat brain. *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
- Cahill, P., Walker, Q. D., Finnegan, J. M., Mickelson, G. E. & Wightman, R. M. Microelectrodes for the measurement of catecholamines in biological systems. *Anal. Chem.* **68**, 3180–3186 (1996).
- Revec, G. V., Grabner, C. P., Pierce, R. C. & Bardo, M. T. Transient increases in catecholaminergic activity in medial prefrontal cortex and nucleus accumbens shell during novelty. *Neuroscience* **76**, 707–714 (1997).
- Schulz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Stellar, J. Jr & Rice, M. B. in *The Neuropharmacological Basis of Reward* (eds Lieberman, J. M. & Cooper, S. J.) 14–65 (Clarendon, Oxford, 1989).
- Fibiger, H. C., LePiane, F. G., Jakubovic, A. & Phillips, A. G. The role of dopamine in intracranial self-stimulation of the ventral tegmental area. *J. Neurosci.* **7**, 3888–3896 (1987).
- Blaha, C. D. & Phillips, A. G. Application of *in vivo* electrochemistry to the measurement of changes in dopamine release during intracranial self-stimulation. *J. Neurosci. Meth.* **34**, 125–133 (1990).
- Miliaressis, A., Emond, C. & Merali, Z. Re-evaluation of the role of dopamine in intracranial self-stimulation using *in vivo* microdialysis. *Behav. Brain Res.* **46**, 43–48 (1991).
- Nakahara, D., Fuchikami, K., Ozaki, N., Iwasaki, T. & Nagatu, T. Differential effect of self-stimulation on dopamine release and metabolism in the rat medial frontal cortex, nucleus accumbens and striatum studied by *in vivo* microdialysis. *Brain Res.* **574**, 164–170 (1992).
- Diorino, D. F., Courty, A., Fibiger, H. C. & Phillips, A. C. Electrical stimulation of reward sites in the ventral tegmental area increases dopamine transmission in the nucleus accumbens of the rat. *Behav. Brain Res.* **55**, 131–141 (1993).
- Murray, B. & Shizgal, P. Evidence implicating both slow- and fast-conducting fibers in the rewarding effect of medial forebrain bundle stimulation. *Behav. Brain Res.* **63**, 47–60 (1994).
- Garris, P. A., Christensen, J. R. C., Rebec, G. V. & Wightman, R. M. Real-time measurement of electrically evoked extracellular dopamine in the striatum of freely moving rats. *J. Neurochem.* **68**, 152–161 (1997).
- Michael, D., Travis, E. R. & Wightman, R. M. Color images for fast-scan CV measurements in biological systems. *Anal. Chem.* **70**, 586A–592A (1998).
- Marsden, C. A. et al. *In vivo* voltammetry—present electrodes and methods. *Neuroscience* **25**, 389–400 (1988).
- Kruk, Z. L. et al. Real time measurement of stimulated dopamine release in the conscious rat using fast cyclic voltammetry: dopamine release is not observed during intracranial self stimulation. *J. Neurosci. Meth.* **79**, 9–19 (1998).
- Jones, S. R., O'Dell, S. J., Marshall, J. F. & Wightman, R. M. Functional and anatomical evidence for different dopamine dynamics in the core and shell of the nucleus accumbens in slices of rat brain. *Synapse* **23**, 224–231 (1996).
- Parsons, L. H. & Justice, J. B. Extracellular concentration and *in vivo* recovery of dopamine in the nucleus accumbens using microdialysis. *J. Neurochem.* **58**, 212–218 (1992).
- Garris, P. A., Ciolkowski, E. L., Pastore, P. & Wightman, R. M. Efflux of dopamine from the synaptic cleft in the nucleus accumbens of the rat brain. *J. Neurosci.* **14**, 6084–6093 (1994).
- Nirenberg, M. J. et al. The dopamine transporter: comparative ultrastructure of dopaminergic axons in limbic and motor compartments of the nucleus accumbens. *J. Neurosci.* **17**, 6899–6907 (1997).
- Michael, A. C., Ikeda, M. & Justice, J. B. Jr Mechanisms contributing to the recovery of striatal releasable dopamine following MFB stimulation. *Brain Res.* **421**, 325–335 (1987).
- Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).
- Mireniewicz, J. & Schultz, W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* **379**, 449–451 (1996).
- Rocha, B. A. et al. Cocaine self-administration in dopamine-transporter knockout mice. *Nature Neurosci.* **1**, 132–137 (1998).
- Gallistel, C. R. & Freyd, G. Quantitative determination of the effects of catecholaminergic agonists and antagonists on the rewarding efficacy of brain stimulation. *Pharm. Biochem. Behav.* **26**, 731–741 (1987).
- Wise, R. A. in *The Neurobiology of Opiate Reward Processes* (eds Smith, J. E. & Lane, J. D.) 405–436 (Elsevier Biomedical, New York, 1983).

**Acknowledgements.** Discussions with G. Breese and H. Sampson are gratefully acknowledged. This research was supported by grants from the NIH (NIDA and NINDS) to R.M.W. and from Illinois State University to P.A.G.

Correspondence and requests for materials should be addressed to R.M.W. (e-mail: rmw@net.chem.unc.edu).

# Regulation of calcium signalling in T lymphocytes by the second messenger cyclic ADP-ribose

Andreas H. Guse\*, Cristina P. da Silva\*, Ingeborg Berg\*, Alla L. Skapenko†, Karin Weber\*, Petra Heyer\*, Martin Hohenegger‡, Gloria A. Ashamu§, Hendrik Schulze-Koops†, Barry V. L. Potter† & Georg W. Mayr\*

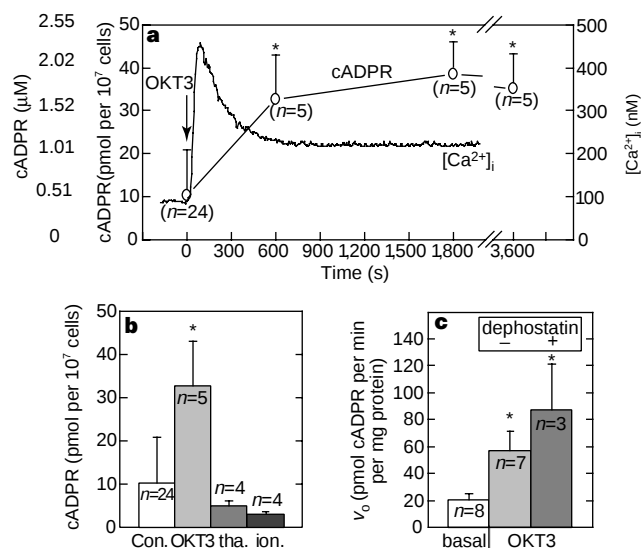
\* Department of Enzyme Chemistry, Institute of Physiological Chemistry, University of Hamburg, Grindelallee 117, 20146 Hamburg, Germany

† Department of Internal Medicine III, University of Erlangen-Nürnberg, 91054 Erlangen, Germany

‡ Institute of Pharmacology, University of Vienna, Währingerstrasse 13A, 1090 Vienna, Austria

§ Wolfson Laboratory for Medicinal Chemistry, Department of Pharmacy and Pharmacology, University of Bath, Claverton Down, Bath BA2 7AY, UK

Cyclic ADP-ribose (cADPR) is a natural compound that mobilizes calcium ions in several eukaryotic cells<sup>1–3</sup>. Although it can lead to the release of calcium ions in T lymphocytes<sup>4–7</sup>, it has not been firmly established as a second messenger in these cells. Here, using high-performance liquid chromatography analysis<sup>8</sup>, we show that stimulation of the T-cell receptor/CD3 (TCR/CD3) complex results in activation of a soluble ADP-ribosyl cyclase and a sustained increase in intracellular levels of cADPR. There is a causal relation between increased cADPR concentrations, sustained calcium signalling and activation of T cells, as shown by

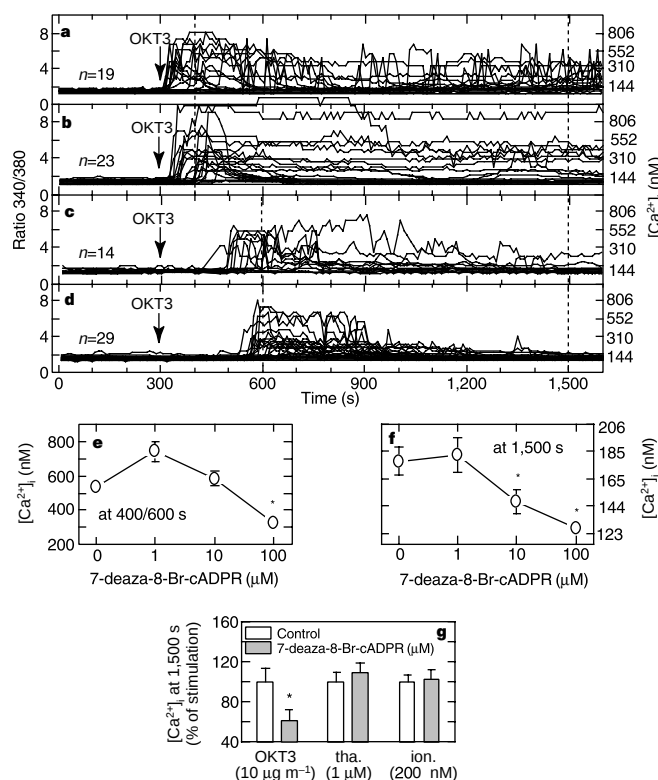


**Figure 1** Endogenous cADPR, ADP-ribosyl cyclase activation and [Ca<sup>2+</sup>]<sub>i</sub> in response to stimulation of the TCR/CD3 complex. **a**, Jurkat T-lymphocytes were left unstimulated for at least 10 min at 37 °C and were then stimulated for the times indicated using 10 μg ml<sup>-1</sup> anti-CD3 OKT3. Analysis of cADPR was carried out as described<sup>8</sup>. [Ca<sup>2+</sup>]<sub>i</sub> was determined in a suspension of fura-2-loaded Jurkat T cells. **b**, Cells were also stimulated by 1 μM thapsigargin (tha.) or 200 nM ionomycin (ion.) for 10 min, and endogenous cADPR was determined. Con., control. **c**, ADP-ribosyl cyclase activity towards NAD<sup>+</sup> was determined in S100 extracts from intact cells that were either unstimulated, stimulated by OKT 3 for 10 min, or stimulated by OKT 3 for 10 min and cyclase prepared in the presence of dephostatin. Data are mean ± s.d. (\*P < 0.05).

inhibition of TCR/CD3-stimulated calcium signalling, cell proliferation and expression of the early- and late-activation markers CD25 and HLA-DR by using cADPR antagonists<sup>9</sup>. The molecular target for cADPR, the type-3 ryanodine receptor/calcium channel, is expressed in T cells. Increased cADPR significantly and specifically stimulates the apparent association of [<sup>3</sup>H]ryanodine with the type-3 ryanodine receptor, indicating a direct modulatory effect of cADPR on channel opening. Thus we show the presence, causal relation and biological significance of the major constituents of the cADPR/calcium-signalling pathway in human T cells.

cADPR mobilizes calcium ions in sea-urchin eggs<sup>1,2,10</sup>, plants<sup>3</sup> and higher eukaryotes (reviewed in ref. 2). In human Jurkat T lymphocytes, cADPR releases Ca<sup>2+</sup> from intracellular stores<sup>4,5</sup>, mediates the entry of Ca<sup>2+</sup> in intact cells<sup>6</sup>, and acts as an endogenous nucleotide<sup>8</sup>. However, it has remained unclear whether cADPR would act as a second messenger in response to stimulation of the TCR/CD3 complex. Previous studies have shown that activation of this complex involves increases in free cytosolic Ca<sup>2+</sup> concentration ([Ca<sup>2+</sup>]<sub>i</sub>). This increase is brought about (1) by release of free cytosolic Ca<sup>2+</sup> through the type 1 inositol 1,4,5-trisphosphate (Ins(1,4,5)P<sub>3</sub>) receptor from intracellular stores, a step known to be essential for CD3-mediated Ca<sup>2+</sup>-signalling<sup>11</sup>; and (2) sustained entry of Ca<sup>2+</sup> (reviewed in ref. 12). Experimental data from T cells<sup>13–15</sup> favour the capacitative model for Ca<sup>2+</sup> entry<sup>16</sup>.

Here we present evidence that cADPR is essential for long-lasting Ca<sup>2+</sup> entry, which is important for the generation of a functional



**Figure 2** Inhibition of long-lasting Ca<sup>2+</sup> signalling by 7-deaza-8-Br-cADPR. [Ca<sup>2+</sup>]<sub>i</sub> was measured in fura-2-loaded Jurkat T cells using single-cell Ca<sup>2+</sup> imaging<sup>6</sup>. Cells were either untreated (**a**) or treated with increasing concentrations of 7-deaza-8-Br-cADPR for 20 min (**b**, 1 μM; **c**, 10 μM; **d**, 100 μM) before stimulation with anti-CD3 OKT3 (10 μg ml<sup>-1</sup>). **e**, **f**, [Ca<sup>2+</sup>]<sub>i</sub> at 400 and 600 s (**e**) or 1,500 s (**f**) after OKT3 addition as combined data obtained in five independent experiments (means ± s.e.m., n = 108–203 cells for each concentration of 7-deaza-8-Br-cADPR. \*P < 0.05). **g**, Effect of 7-deaza-8-Br-cADPR on cells stimulated by either OKT3, thapsigargin or ionomycin (mean ± s.d., 3–4 independent experiments, 91–104 individual cells investigated. \*P < 0.05).

immune response because T cells require the sustained  $\text{Ca}^{2+}$  signal for clonal expansion. We found that the agonistic anti-CD3 antibody OKT3 induced a slowly rising but sustained increase in intracellular cADPR, as measured by high-performance liquid chromatography, reaching its highest level 30 min after stimulation; this level was still maintained after 60 min (Fig. 1a). The concentration of intracellular inorganic phosphate ( $\text{P}_i$ ), which modulates the effects of cADPR by increasing the loading of cADPR-sensitive  $\text{Ca}^{2+}$  stores<sup>5</sup>, may also contribute to the action of cADPR.

In contrast to OKT3, neither stimulating  $\text{Ca}^{2+}$  signalling by artificial store depletion using thapsigargin nor increasing  $[\text{Ca}^{2+}]_i$  by using a moderate concentration of ionomycin increased the level of endogenous cADPR (Fig. 1b). The ADP-ribosyl cyclase involved is therefore not stimulated unspecifically by increased  $[\text{Ca}^{2+}]_i$ .

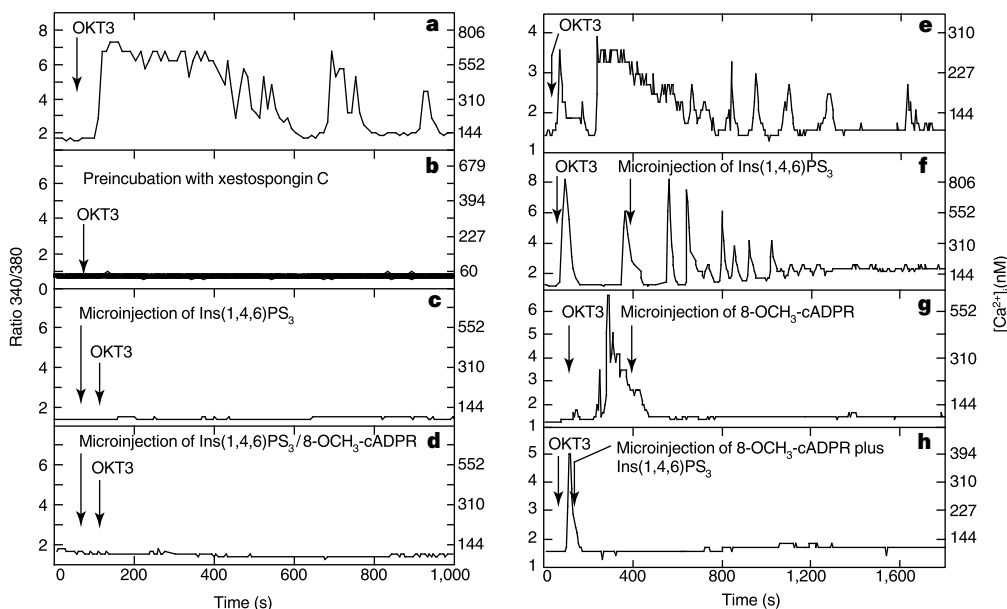
As a potential link between TCR/CD3 stimulation and increased cADPR concentration, we detected a new ADP-ribosyl cyclase at low basal activity in the cytosolic fraction of Jurkat cells (Fig. 1c). In contrast to known ADP-ribosyl cyclases, such as the soluble one from *Aplysia californica*<sup>17</sup> or the plasma-membrane-bound ectoenzyme CD38 (ref. 18), the soluble enzyme from Jurkat T cells did not use as a substrate nicotinamide guaninedinucleotide (data not shown). Most importantly, the initial reaction velocity of this enzyme was increased significantly in T cells stimulated by OKT3 (Fig. 1c). This effect was greater when the cytosolic fraction was prepared in the presence of the Tyr-phosphatase inhibitor dephostatin<sup>19</sup> (Fig. 1c), indicating activation by Tyr-phosphorylation of the soluble cyclase.

In order to demonstrate a causal relation between the TCR/CD3-mediated increase in cADPR level and  $[\text{Ca}^{2+}]_i$ , we preincubated Jurkat T cells with the membrane-permeant cADPR antagonist 7-deaza-8-Br-cADPR<sup>9</sup> and measured OKT3-stimulated calcium signalling (Fig. 2). Most of the cells exhibited  $\text{Ca}^{2+}$  signalling for more than 20 min after stimulation (Fig. 2a). Such a response was not observed in the absence of extracellular  $\text{Ca}^{2+}$  (data not shown), indicating that the signalling was brought about by  $\text{Ca}^{2+}$  entry. Preincubation with 7-deaza-8-Br-cADPR significantly inhibited the long-lasting  $\text{Ca}^{2+}$  entry (Fig. 2b–d, f), indicating that cADPR is required for  $\text{Ca}^{2+}$  entry. In addition, high concentrations of 7-deaza-

8-Br-cADPR partly inhibited the rapid phase of  $\text{Ca}^{2+}$  signalling (Fig. 2e) and increased the delay between addition of OKT3 and the onset of the  $\text{Ca}^{2+}$  signal (Fig. 2c, d), indicating that cADPR is involved in the early phase of  $\text{Ca}^{2+}$  signalling. It is unlikely that 7-deaza-8-Br-cADPR has an unspecific effect on  $\text{Ins}(1,4,5)\text{P}_3$ -induced  $\text{Ca}^{2+}$  release because the closely related derivative 8-Br-cADPR, even at concentrations up to 100  $\mu\text{M}$ , has been shown not to inhibit  $\text{Ins}(1,4,5)\text{P}_3$ -induced  $\text{Ca}^{2+}$  release in T cells<sup>4</sup>. Importantly, 7-deaza-8-Br-cADPR did not inhibit the sustained  $\text{Ca}^{2+}$  signalling stimulated by either store depletion using thapsigargin or by unspecific increase of  $[\text{Ca}^{2+}]_i$  by ionomycin (Fig. 2g).

We found that 7-deaza-8-Br-cADPR only partly inhibited the first phase of  $\text{Ca}^{2+}$  signalling (Fig. 2), which is thought to be driven mainly by  $\text{Ins}(1,4,5)\text{P}_3$ . The membrane-permeant  $\text{Ins}(1,4,5)\text{P}_3$  antagonist xestospongine C<sup>20</sup> has been shown to potentially block  $\text{Ins}(1,4,5)\text{P}_3$ -mediated  $\text{Ca}^{2+}$  release but, up to extracellular concentration of 20  $\mu\text{M}$ , has no effect on ryanodine receptor (RyR)-mediated  $\text{Ca}^{2+}$  release<sup>20</sup>. We found that xestospongine C inhibits TCR/CD3-mediated  $\text{Ca}^{2+}$  signalling after preincubation of the cells for 20 min at 2  $\mu\text{M}$  (Fig. 3a, b). Similarly, microinjection of single T cells with either the  $\text{Ins}(1,4,5)\text{P}_3$  antagonist inositol 1,4,6-phosphorothioate ( $\text{Ins}(1,4,6)\text{PS}_3$ ) or a combined injection of  $\text{Ins}(1,4,6)\text{PS}_3$  and the cADPR antagonist 8-methoxy-cADPR, before stimulation with OKT3, completely abolished TCR/CD3-mediated  $\text{Ca}^{2+}$  signalling (Fig. 3c, d). In contrast, microinjection of  $\text{Ins}(1,4,6)\text{PS}_3$  after the development of OKT3-stimulated  $\text{Ca}^{2+}$  signalling was almost without effect (Fig. 3f), whereas such late injection of 8-methoxy-cADPR, either alone or in combination with  $\text{Ins}(1,4,6)\text{PS}_3$ , completely inhibited further  $\text{Ca}^{2+}$  signalling (Fig. 3g, h). These results, taken together with the data from the type 1  $\text{Ins}(1,4,5)\text{P}_3$ -receptor knockout T-cell line<sup>11</sup>, indicate: (1) that an initial  $\text{Ins}(1,4,5)\text{P}_3$ -mediated  $\text{Ca}^{2+}$  signalling event is necessary but not sufficient for sustained  $\text{Ca}^{2+}$  signalling; (2) that cADPR is essential for the sustained second phase; and (3) that cADPR may also be involved in determining the time of onset of  $\text{Ca}^{2+}$  signals after TCR/CD3 stimulation.

$\text{Ca}^{2+}$  release by cADPR has been proposed to proceed by means of



**Figure 3** Effect of antagonists against  $\text{Ins}(1,4,5)\text{P}_3$  and cADPR on  $\text{Ca}^{2+}$  signalling in single Jurkat T cells.  $[\text{Ca}^{2+}]_i$  was measured as above. T cells were stimulated by OKT3 and either left untreated (a, e,  $n > 100$ ) or preincubated with xestospongine C (2  $\mu\text{M}$ , 20 min) before addition of OKT3 (b,  $n = 16$ ), microinjected<sup>6</sup> with 80  $\mu\text{M}$   $\text{Ins}(1,4,6)\text{PS}_3$  (c,  $n = 3$ ) or with 80  $\mu\text{M}$   $\text{Ins}(1,4,6)\text{PS}_3$  plus 100  $\mu\text{M}$  8-methoxy cADPR

before addition of OKT3 (d,  $n = 3$ ) or microinjected after the first OKT3-dependent  $\text{Ca}^{2+}$  spikes were observed with 80  $\mu\text{M}$   $\text{Ins}(1,4,6)\text{PS}_3$  (f,  $n = 10$ ), 100  $\mu\text{M}$  8-methoxy cADPR (g,  $n = 11$ ) or a combination of both (h,  $n = 12$ ). In b, 16 cells monitored in one experiment are displayed as an overlay.



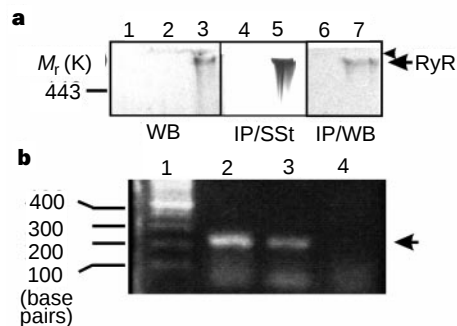
ryanodine receptors<sup>10</sup>. We have obtained pharmacological evidence that cADPR acts through ryanodine receptors in T cells<sup>4,5</sup>. Despite these results, and the fact that the  $\text{Ca}^{2+}$ -mobilizing effects of cADPR in T cells have been reported independently several times<sup>4,7,21</sup>, the expression of ryanodine receptors in T cells has been a matter of debate. They have been detected in T cells at the DNA, RNA and protein level<sup>7,22,23</sup>. In contrast, transcripts for ryanodine receptors have not been detected using the polymerase chain reaction with reverse transcriptase (RT-PCR) in another Jurkat clone<sup>24</sup>. Here we show that ryanodine receptors are in our Jurkat T-cell line, as shown by western blot and immunoprecipitation experiments using an anti-ryanodine-receptor monoclonal antibody (anti-RyR<sub>common</sub>) (Fig. 4a). Ryanodine receptors were detected in P10 membranes prepared from Jurkat cells, but not in the cytosolic fraction (Fig. 4a). Furthermore, immunoprecipitation of solubilized P10 membranes with the anti-RyR<sub>common</sub> monoclonal antibody resulted in profound enrichment of the ryanodine receptor (Fig. 4a). These data were confirmed by RT-PCR of poly(A)<sup>+</sup> RNA from Jurkat T cells. Using redundant primers for all ryanodine-receptor subtypes, a PCR product of the expected size was obtained. The nucleotide sequence from the amplified product was identical to a sequence from type 3 ryanodine receptor. A sequence identical to a type 3 ryanodine-receptor sequence was also obtained when specific primers were used (Fig. 4b).

We also observed specific binding of [<sup>3</sup>H]ryanodine to Jurkat membranes. cADPR was found to have a significant stimulatory effect on the apparent association velocity of [<sup>3</sup>H]ryanodine, but not on the capacity of [<sup>3</sup>H]ryanodine binding to Jurkat membranes (Fig. 5a). These findings indicate a direct modulatory effect on channel opening by cADPR, as [<sup>3</sup>H]ryanodine has been suggested to bind the ryanodine receptor in the open conformation<sup>25,26</sup>. A similar kinetic effect of cADPR has been reported for  $\text{Ca}^{2+}$  release from rat cardiac and brain microsomes<sup>27</sup>. The stimulatory effect of cADPR on the binding velocity of ryanodine was dose dependent (Fig. 5b) and specific, as it was not observed with ATP or NAD (Fig. 5b), and it was reversed by various antagonists of ryanodine receptor and of cADPR (Fig. 5c). High concentrations of  $\text{Mg}^{2+}$  and ruthenium red decreased [<sup>3</sup>H]ryanodine binding below control levels, whereas the cADPR antagonists reversed only the stimulatory effect of cADPR (Fig. 5c).

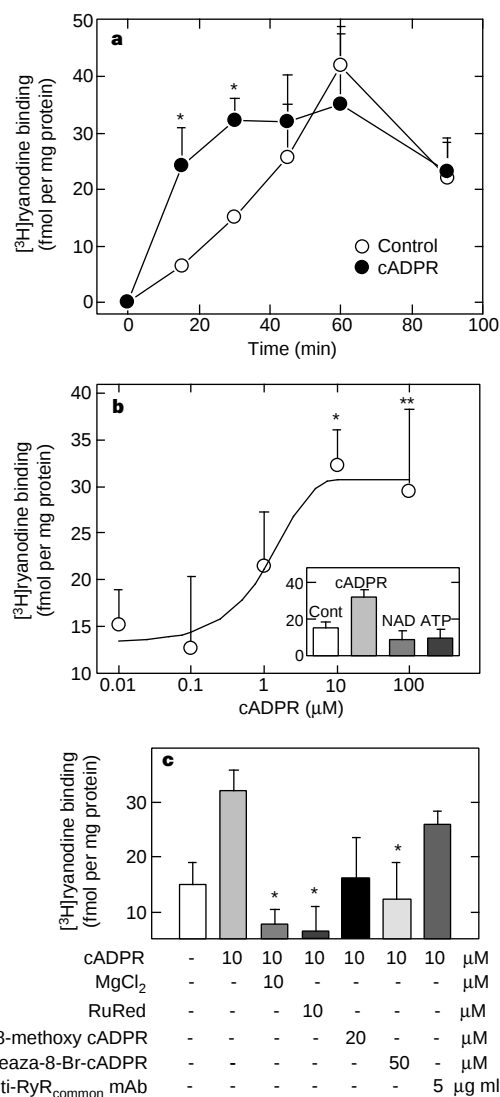
To confirm that cADPR is required for the activation of peripheral T cells, these cells were activated by solid-phase-bound

anti-CD3 monoclonal antibody in the presence of increasing concentrations of 7-deaza-8-Br-cADPR. Proliferation was inhibited partially by 10  $\mu\text{M}$  and completely by 100  $\mu\text{M}$  7-deaza-8-Br-cADPR (Table 1). Similarly, the expression of the T-cell activation markers CD25 and HLA-DR was markedly suppressed by 7-deaza-8-Br-cADPR (Table 1). Most importantly, after five days of exposure to 100  $\mu\text{M}$  7-deaza-8-Br-cADPR, more than 70% of the cells still were detected in the lymphocyte gate, of which more than 95% were negative for propidium iodine staining, indicating that the antagonist has almost no unspecific cytotoxic effect.

We propose that cADPR is formed as an essential second messenger in response to stimulation of the TCR/CD3 complex, and mediates the second, sustained phase of  $\text{Ca}^{2+}$  signalling by means of the ryanodine receptor/ $\text{Ca}^{2+}$  channel in T cells. Pancreatic acinar cells appear to respond to agonist stimulation by either the  $\text{Ins}(1,4,5)\text{P}_3$  or the cADPR pathway; this seems to be controlled by the intracellular glucose concentration<sup>28</sup>. However, because initial



**Figure 4** Expression of ryanodine receptor in Jurkat T-lymphocytes. **a**, T-cell fractions or immunoprecipitated ryanodine receptor (lane 1, S100; 2, p100; 3, P10; each 100  $\mu\text{g}$  protein; 4 and 6, IP supernatant, 10  $\mu\text{g}$ ; 5 and 7, immunoprecipitate 1  $\mu\text{g}$ ) were subjected to reducing SDS-PAGE, either transferred to nitrocellulose sheets and immunostained (lanes 1-3, 6, 7) or silver stained (lanes 4, 5). Arrow, ryanodine receptor (RyR); arrowhead, top of separation gel; WB, western blot; IP, immunoprecipitation; SSt, silver staining). **b**, Agarose gel electrophoresis of PCR products (from two preparations of Jurkat T-cell mRNA, lanes 2 and 3) using specific primers for type 3 ryanodine receptor (lane 1, relative molecular mass ( $M_r$ ) marker; lane 4, water control).



**Figure 5** Effect of cADPR on [<sup>3</sup>H]ryanodine binding to P10 membranes. **a**, Kinetics of the specific binding of 200 nM [<sup>3</sup>H]ryanodine in the absence or presence of cADPR (10  $\mu\text{M}$ ;  $n = 4-8$ ). **b**, Dose dependency and specificity of the effect of cADPR on the velocity of specific [<sup>3</sup>H]ryanodine binding, assessed by a 30-min incubation of membranes at 37 °C in the presence of cADPR, or for control in the presence of NAD (10  $\mu\text{M}$ ) or ATP (10  $\mu\text{M}$ ;  $n = 4-8$ ; \*\* $P < 0.1$ ; \* $P < 0.05$ ). **c**, Effect of various compounds on cADPR-mediated specific binding of 200 nM [<sup>3</sup>H]ryanodine to P10 membranes (30 min, 37 °C,  $n = 3-8$ ; \* $P < 0.05$ ).

**Table 1 Inhibition of T-cell activation by 7-deaza-8-Br-cADPR**

| Addition                                   | [ <sup>3</sup> H]thymidine incorporation<br>[c.p.m. ± s.e.m.] | Expression of<br>CD25 (%) | Expression of<br>HLA-DR (%) |
|--------------------------------------------|---------------------------------------------------------------|---------------------------|-----------------------------|
| Control                                    | 148,494 ± 17,414                                              | 24.5                      | 6.0                         |
| 7-deaza-8-Br-cADPR<br>(1 μM)               | 156,730 ± 13,335                                              | 16.1                      | 7.6                         |
| 7-deaza-8-Br-cADPR<br>(10 μM)              | 94,124 ± 9,020                                                | 14.1                      | 5.3                         |
| 7-deaza-8-Br-cADPR<br>(100 μM)             | 42 ± 8                                                        | 9.8                       | 2.7                         |
| Cyclosporin A<br>(50 ng ml <sup>-1</sup> ) | 41,807 ± 16,424                                               | 12.7                      | 4.9                         |

Inhibitors were added to peripheral blood T cells for 30 min in the absence of serum. The cells were stimulated by solid-phase bound OKT3 in full medium. Proliferation, as assessed by [<sup>3</sup>H]thymidine incorporation and expression of activation markers, determined by FACS analysis, were studied. Results are for quadruplicate determinations from one out of three independent experiments with similar results, using three different healthy donors. Cyclosporin A was added for comparison as an established inhibitor of T-cell activation.

blockade of Ins(1,4,5)P<sub>3</sub>-dependent Ca<sup>2+</sup> release blocked all further Ca<sup>2+</sup> signalling in T cells, the two second-messenger systems seem to act in temporal sequence rather than in an alternative manner. Although we have delineated the signalling pathway from the TCR/CD3 complex in the plasma membrane up to the ryanodine receptor, the exact mechanism by which cADPR regulates Ca<sup>2+</sup> entry is not yet clear. However, because cADPR also mediated depletion at the Ca<sup>2+</sup> pool by releasing Ca<sup>2+</sup> from its target store, it might act in a similar way to that suggested for Ins(1,4,5)P<sub>3</sub> in the classical capacitative model<sup>16</sup>. □

## Methods

**Preparation of subcellular fractions.** Jurkat T cells (10<sup>9</sup> cells) were disrupted in 5 ml 20 mM HEPES (pH 7.5), 110 mM NaCl and protease inhibitors (antipain 5 μg ml<sup>-1</sup>, leupeptin 5 μg ml<sup>-1</sup>, pepstatin 6.9 μg ml<sup>-1</sup>, Pefablock SC 13.9 μg ml<sup>-1</sup>) in a Potter–Elvehjem homogenizer at 4 °C. Cell debris was removed at 500g for 10 min, and the supernatant was further centrifuged at 10,000g for 20 min at 4 °C before the pellet (P10) was collected. The supernatant was then centrifuged at 100,000g for 120 min at 4 °C, and the pellet (P100) and the supernatant (S100) were collected.

**Assay for ADP-ribosyl cyclase.** To determine ADP-ribosyl cyclase activity, NAD<sup>+</sup> (100 μM) was incubated with cytosolic (S100) protein (2 mg ml<sup>-1</sup>) at 37 °C for 0, 1, 2, 3, 5 and 10 min. The reaction was stopped by addition of 3 M perchloric acid, and cADPR was analysed by reverse-phase HPLC as described<sup>8</sup>. S100 protein was obtained from unstimulated cells (basal activity) or from cells stimulated by OKT3 (10 μg ml<sup>-1</sup>) for 10 min; in some experiments, dephostatin (100 μM) was present during preparation of S100 protein.

**Immunoprecipitation, SDS-PAGE and western blotting.** Anti-RyR<sub>common</sub> monoclonal antibody was coupled to G-Sepharose and incubated with solubilized P10 membranes in lysis buffer (25 mM HEPES, pH 7.2, 150 mM NaCl, 0.25% CHAPS (w/v) and protease inhibitors as above) for 1.5 h at 4 °C under continuous shaking. The G-Sepharose beads were spun down at 13,000g for 1 min at 4 °C, the supernatant was discarded, and the immunoprecipitate rinsed with lysis buffer; this procedure was then repeated three times. Protein was boiled at 95 °C for 5 min and subjected to SDS-PAGE in a 6% gel (3% stacking gel) under reducing conditions. Proteins were then transferred onto nitrocellulose sheets by tank blotting (18 h, 550 mA constant, 4 °C). The nitrocellulose sheets were immunostained with anti-RyR<sub>common</sub> monoclonal antibody (5 μg ml<sup>-1</sup>, 24 h), washed by a complex procedure and developed using the ECL kit (Amersham) according to the manufacturer's instructions.

**Binding of [<sup>3</sup>H]ryanodine to P10 membranes.** Binding of [<sup>3</sup>H]ryanodine was measured in 20 mM HEPES, pH 7.4, KCl 0.75 M, CaCl<sub>2</sub> 1.1 mM, EGTA 1.0 mM ([Ca<sup>2+</sup>]<sub>free</sub> ≈ 100 μM) using 200 μg protein in a 50-μl volume at 37 °C with continuous shaking. Unspecific binding was determined in the presence of a 1,000-fold excess of unlabelled ryanodine. Separation of bound from free radioactivity was achieved by rapid filtration on glass-fibre filters (Whatman type GF/B). The filters were rinsed rapidly four times with 3 ml ice-cold buffer containing 10 mM HEPES, pH 7.4, KCl 0.75 M, CaCl<sub>2</sub> 1.1 mM, EGTA 1.0 mM

([Ca<sup>2+</sup>]<sub>free</sub> ≈ 100 μM), and subjected to liquid scintillation counting. Analysis of binding isotherms (at 37 °C) resulted in a K<sub>d</sub> of 43 ± 18 nM (n = 3) and a maximal number of specific binding sites of 66 ± 11 fmol per mg protein (n = 3).

**RT-PCR.** To amplify type 3 ryanodine-receptor complementary DNA sequences, the following specific primers were used: sense, 5'-CGACATGATGACGTGTACC-3'; antisense, 5'-CTCGTACTGTCTCTGCTGG-3').

**Other procedures.** HPLC analysis of cADPR, radiometric digital Ca<sup>2+</sup>-imaging, microinjection, FACS analysis and proliferation assay for peripheral T cells were carried out as described<sup>6,8,29,30</sup>. See Supplementary Information for further details.

Received 14 December 1998; accepted 19 January 1999.

- Lee, H. C., Walseth, T. F., Bratt, G. T., Hayes, R. H. & Clapper, D. L. Structural determination of a cyclic metabolite of NAD<sup>+</sup> with intracellular Ca<sup>2+</sup> mobilizing activity. *J. Biol. Chem.* **264**, 1608–1615 (1989).
- Lee, H. C. Mechanisms of calcium signaling by cyclic ADP-ribose and NAADP. *Physiol. Rev.* **77**, 1133–1164 (1997).
- Wu, Y. *et al.* Abscise and signaling through cyclic ADP-ribose in plants. *Science* **278**, 2126–2130 (1997).
- Guse, A. H. *et al.* Characterization of cyclic adenosine diphosphate-ribose-induced Ca<sup>2+</sup>-release in T-lymphocyte cell lines. *J. Immunol.* **155**, 3353–3359 (1995).
- Guse, A. H. *et al.* Regulation of cyclic ADP-ribose-induced Ca<sup>2+</sup>-release by Mg<sup>2+</sup> and inorganic phosphate. *J. Biol. Chem.* **271**, 23946–23954 (1996).
- Guse, A. H., Berg, L., da Silva, C. P., Potter, B. V. L. & Mayr, G. W. Ca<sup>2+</sup>-entry induced by cyclic ADP-ribose in intact T-lymphocytes. *J. Biol. Chem.* **272**, 8546–8550 (1997).
- Bourguignon, L. Y. W., Chu, A., Jin, H. & Brandt, N. R. Ryanodine receptor–ankyrin interaction regulates internal Ca<sup>2+</sup>-release in mouse T-lymphoma cells. *J. Biol. Chem.* **270**, 17917–17922 (1995).
- da Silva, C. P., Potter, B. V. L., Mayr, G. W. & Guse, A. H. Quantification of intracellular levels of cyclic ADP-ribose by high-performance liquid chromatography. *J. Chromatogr. B* **707**, 43–50 (1998).
- Sethi, J. K., Empson, R. M., Bailey, V. C., Potter, B. V. L. & Galione, A. 7-Deaza-8-bromo-cyclic ADP-ribose, the first membrane-permeant, hydrolysis-resistant cyclic ADP-ribose antagonist. *J. Biol. Chem.* **272**, 16358–16363 (1997).
- Galione, A., Lee, H. C. & Busa, W. B. Ca<sup>2+</sup>-induced Ca<sup>2+</sup>-release in sea urchin egg homogenates: modulation by cyclic ADP-ribose. *Science* **253**, 1143–1146 (1991).
- Jayaraman, T., Ondriasova, E., Ondrias, K., Harnick, D. J. & Marks, A. R. The inositol 1,4,5-trisphosphate receptor is essential for T-cell receptor signaling. *Proc. Natl Acad. Sci. USA* **92**, 6007–6011 (1995).
- Guse, A. H. Ca<sup>2+</sup>-signaling in T-lymphocytes. *Crit. Rev. Immunol.* **18**, 419–448 (1988).
- Zweifach, A. & Lewis, R. S. Mitogen-regulated Ca<sup>2+</sup>-current of T lymphocytes is activated by depletion of intracellular Ca<sup>2+</sup>-stores. *Proc. Natl Acad. Sci. USA* **90**, 6295–6299 (1993).
- Partiseti, M. *et al.* The calcium current activated by the T cell receptor and store depletion in human lymphocytes is absent in a primary immunodeficiency. *J. Biol. Chem.* **269**, 32327–32335 (1994).
- Fanger, C. M., Hoth, M., Crabtree, G. R. & Lewis, R. S. Characterization of T cell mutants with defects in capacitative calcium entry: genetic evidence for the physiological roles of CRAC channels. *J. Cell Biol.* **131**, 655–667 (1995).
- Putney, J. W. Jr A model for receptor-regulated calcium entry. *Cell Calcium* **7**, 1–12 (1986).
- Hellmich, M. R. & Strumwasser, F. Purification and characterization of a molluscan egg-specific NADase, a second messenger enzyme. *Cell Regul.* **2**, 193–202 (1991).
- Howard, M. *et al.* Formation and hydrolysis of cyclic ADP-ribose catalyzed by lymphocyte antigen CD38. *Science* **262**, 1056–1059 (1993).
- Imoto, M. *et al.* Dephostatin, a novel protein tyrosine phosphatase inhibitor produced by *Streptomyces*. I. Taxonomy, isolation and characterization. *J. Antibiot.* **46**, 1342–1346 (1993).
- Gafni, J. *et al.* Xestospingins: potent membrane permeable blockers of the inositol 1,4,5-trisphosphate receptor. *Neuron* **19**, 723–733 (1997).
- Takemura, H., Imoto, K., Sakano, S., Kaneko, M. & Ohshika, H. Lysophosphatidic acid-sensitive intracellular Ca<sup>2+</sup> store does not regulate Ca<sup>2+</sup> entry at plasma membrane in Jurkat human T-cells. *Biochem. J.* **319**, 393–397 (1996).
- Ricard, I., Martel, J., Dupuis, L., Dupuis, G. & Payet, M. D. A caffeine/ryanodine-sensitive Ca<sup>2+</sup> pool is involved in triggering spontaneous variations of Ca<sup>2+</sup> in Jurkat T-lymphocytes by a Ca<sup>2+</sup>-induced Ca<sup>2+</sup> release (CICR) mechanism. *Cell. Signal.* **9**, 197–206 (1997).
- Hakamata, Y. *et al.* Involvement of the brain type of ryanodine receptor in T-cell proliferation. *FEBS Lett.* **352**, 206–210 (1994).
- Bennett, D. L. *et al.* Expression and function of ryanodine receptors in nonexcitable cells. *J. Biol. Chem.* **271**, 6356–6362 (1996).
- Meissner, G. Ryanodine receptor/Ca<sup>2+</sup> release channels and their regulation by endogenous effectors. *Annu. Rev. Physiol.* **56**, 485–508 (1994).
- Sutko, J. L. & Airey, J. A. Ryanodine receptor Ca<sup>2+</sup> release channels: does diversity in form equal diversity in function? *Physiol. Rev.* **76**, 1027–1071 (1996).
- Mészáros, L. G., Bak, J. & Chu, A. Cyclic ADP-ribose as an endogenous regulator of the non-skeletal type ryanodine receptor Ca<sup>2+</sup> channel. *Nature* **364**, 76–79 (1993).
- Cancela, J. M., Mogami, H., Tepikin, A. V. & Petersen, O. H. Intracellular glucose switches between cyclic ADP-ribose and inositol trisphosphate triggering of cytosolic Ca<sup>2+</sup>-spiking. *Curr. Biol.* **16**, 865–868 (1998).
- Schulze-Koops, H., Lipsky, P. E., Kavanaugh, A. F. & Davis, L. S. Elevated Th1- or Th0-like cytokine mRNA in peripheral circulation of patients with rheumatoid arthritis: modulation by treatment with anti-ICAM-1 correlates with clinical benefit. *J. Immunol.* **155**, 5029–5037 (1995).
- Schulze-Koops, H., Lipsky, P. E. & Davis, L. S. Human memory T-cell differentiation into Th2-like effector cells is dependent on IL-4 and CD28 stimulation and inhibited by TCR ligation. *Eur. J. Immunol.* **28**, 2517–2529 (1998).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank M. K. Jacobson and K.-H. Krause for comments on the manuscript, and V. C. Bailey and S. J. Mills for synthesizing some of the ligands used. This work was supported by the Deutsche Forschungsgemeinschaft (A.H.G., G.W.M. and H.S.-K.), Alexander-von-Humboldt-Stiftung (C.P.S.), Büro zur Förderung von Auslandsbeziehungen und Österreichische Nationalbank (M.H.) and the Wellcome Trust (B.V.L.P. and A.H.G.).

Correspondence and requests for materials should be addressed to A.H.G. (e-mail: [guse@uke.uni-hamburg.de](mailto:guse@uke.uni-hamburg.de))

# Coordination of agonist-induced $\text{Ca}^{2+}$ -signalling patterns by NAADP in pancreatic acinar cells

Jose Manuel Cancela\*, Grant C. Churchill & Antony Galione

Department of Pharmacology, Mansfield Road, University of Oxford, Oxford OX1 3QT, UK

\* Present address: Department of Physiology, University of Liverpool, Crown Street, Liverpool L69 3BX, UK.

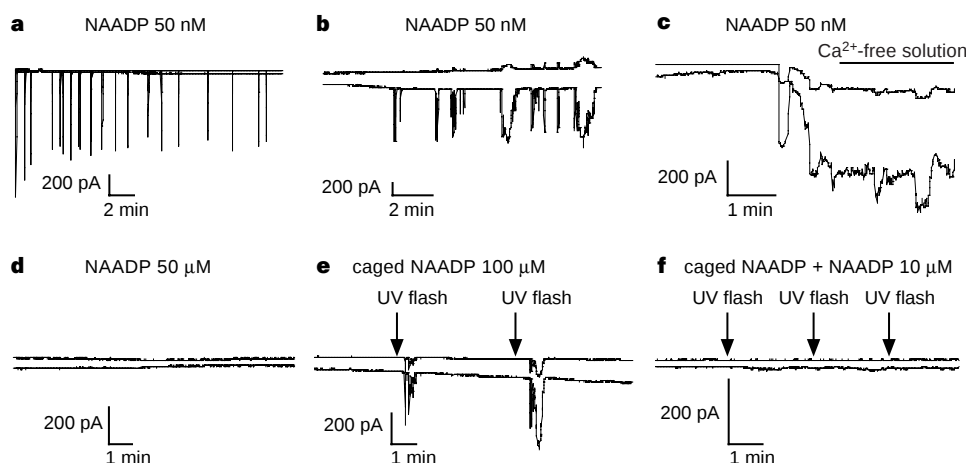
Many hormones and neurotransmitters evoke  $\text{Ca}^{2+}$  release from intracellular stores, often triggering agonist-specific signatures of intracellular  $\text{Ca}^{2+}$  concentration<sup>1–5</sup>. Inositol trisphosphate ( $\text{InsP}_3$ )<sup>1</sup> and cyclic adenosine 5'-diphosphate-ribose ( $\text{cADPR}$ )<sup>6,7</sup> are established  $\text{Ca}^{2+}$ -mobilizing messengers that activate  $\text{Ca}^{2+}$  release through intracellular  $\text{InsP}_3$  and ryanodine receptors, respectively<sup>8–10</sup>. However, in pancreatic acinar cells, neither messenger can explain the complex pattern of  $\text{Ca}^{2+}$  signals triggered by the secretory hormone cholecystokinin (CCK). We show here that the  $\text{Ca}^{2+}$ -mobilizing molecule nicotinic acid adenine dinucleotide phosphate (NAADP)<sup>7,11–15</sup>, an endogenous metabolite of  $\beta$ -NADP, triggers a  $\text{Ca}^{2+}$  response that varies from short-lasting  $\text{Ca}^{2+}$  spikes to a complex mixture of short-lasting (1–2 s) and long-lasting (0.2–1 min)  $\text{Ca}^{2+}$  spikes. Cells were significantly more sensitive to NAADP than to either  $\text{cADPR}$  or  $\text{InsP}_3$ , whereas higher concentrations of NAADP selectively inactivated CCK-evoked  $\text{Ca}^{2+}$  signals in pancreatic acinar cells, indicating that NAADP may function as an intracellular messenger in mammalian cells.

To investigate possible roles in mammalian cells for NAADP, a potent  $\text{Ca}^{2+}$ -mobilizing agent in marine invertebrate eggs<sup>7,11–15</sup>, we examined its effect on  $\text{Ca}^{2+}$  signals in pancreatic acinar cells. Use of the whole-cell configuration of the patch-clamp technique allowed us to gain access to the cytosol to monitor changes in the concentration of cytosolic  $\text{Ca}^{2+}$  by measuring  $\text{Ca}^{2+}$ -dependent ionic currents, which have been extensively correlated with different  $\text{Ca}^{2+}$  signalling patterns<sup>16,17</sup>. When NAADP was infused through the intracellular patch-pipette solution it elicited concentration-

dependent  $\text{Ca}^{2+}$  spiking (Fig. 1). With 5 nM NAADP in the pipette solution, only two out of five cells displayed three  $\text{Ca}^{2+}$  spikes in 20 min and one out of five cells showed a sustained response. The intensity of the response to 50 nM NAADP ( $n = 21/27$ ) varied and we classified it into three categories: short-lasting  $\text{Ca}^{2+}$  spikes (1–2 s) with a frequency of 1  $\text{Ca}^{2+}$  spikes per min (Fig. 1a,  $n = 4/21$ ); complex patterns of  $\text{Ca}^{2+}$  spikes, which are a mixture of short-lasting (1–2 s) and long-lasting (0.2–1 min)  $\text{Ca}^{2+}$  transients (Fig. 1b,  $n = 7/21$ ); and sustained  $\text{Ca}^{2+}$  release ( $n = 10/21$ ). Perfusion (extracellular perfusion) of cells with a  $\text{Ca}^{2+}$ -free solution did not prevent the NAADP-evoked sustained  $\text{Ca}^{2+}$  response ( $n = 3$ ), indicating that  $\text{Ca}^{2+}$  is released from intracellular stores (Fig. 1c). With NAADP in the concentration range 100–500 nM, only one cell out of six responded.

Surprisingly, higher NAADP concentrations (1–100  $\mu\text{M}$ ) in the patch-pipette solution failed to evoke any  $\text{Ca}^{2+}$  release (Fig. 1d,  $n = 19$ ). This may be because, at these concentrations, NAADP diffusion is very fast and desensitizes NAADP receptors before any elevation in intracellular  $\text{Ca}^{2+}$  can be detected. In sea urchin eggs, the NAADP-sensitive  $\text{Ca}^{2+}$  release mechanism is particularly prone to inactivation because even at sub-threshold concentrations of NAADP for  $\text{Ca}^{2+}$  release, NAADP can fully desensitize putative NAADP receptors<sup>15,18,19</sup>. To investigate this inactivation further, we used an inactive caged form of NAADP, 1(2-nitrophenyl)-ethyl-NAADP (ref. 20), which under ultraviolet flash illumination releases the active form, NAADP. With 100  $\mu\text{M}$  of the caged NAADP in the patch-pipette solution, all 11 cells investigated were silent before ultraviolet illumination, showing that the caged compound is inactive. After an ultraviolet flash, the released NAADP (estimated final concentration 0.5–1  $\mu\text{M}$ ) triggered either typical complex patterns of short-lasting and long-lasting  $\text{Ca}^{2+}$  spikes (Fig. 1e,  $n = 5/11$ ) or  $\text{Ca}^{2+}$  oscillations ( $n = 2/11$ , not shown). In the seven responsive cells, a second ultraviolet flash triggered a  $\text{Ca}^{2+}$  release similar to the first  $\text{Ca}^{2+}$  response (Fig. 1e). However, in the presence of a high NAADP concentration (10  $\mu\text{M}$  in the pipette solution), photolysis of caged NAADP (100  $\mu\text{M}$ ) failed to evoke any  $\text{Ca}^{2+}$  release, suggesting that micromolar concentrations of NAADP are required to desensitize the NAADP receptor (Fig. 1f,  $n = 7$ ).

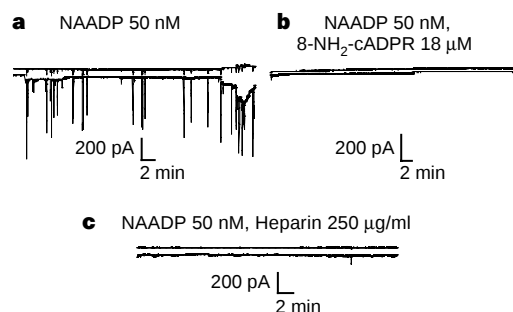
Next, we investigated the characteristics of intracellular  $\text{Ca}^{2+}$  stores mobilized by NAADP. We used 8- $\text{NH}_2$ - $\text{cADPR}$ <sup>21,22</sup> and heparin<sup>23</sup> as selective inhibitors of  $\text{cADPR}$  and  $\text{InsP}_3$ -sensitive  $\text{Ca}^{2+}$  release, respectively. Although these agents fail to affect NAADP-induced  $\text{Ca}^{2+}$  release in sea urchin egg homogenates<sup>11</sup>,



**Figure 1** Intracellular NAADP evokes different patterns of  $\text{Ca}^{2+}$  spikes. The traces represent recordings from single or double cells; NAADP was present in the pipette solution at 50 nM (**a**, **b**, **c**). Trace **b** represents an extract of the recording; complex  $\text{Ca}^{2+}$  signals often occurred after 20 min and when cells were perfused

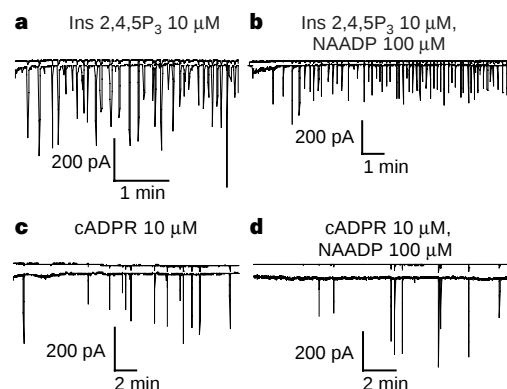
with NAADP at high  $\text{Ca}^{2+}$  resting levels. At concentrations from 1 to 100  $\mu\text{M}$  in the pipette solution (**d**), NAADP failed to evoke  $\text{Ca}^{2+}$  release (NAADP obtained from Sigma was checked for purity by HPLC<sup>11</sup>). The trace with 50  $\mu\text{M}$  NAADP is shown as an example. Caged NAADP at 100  $\mu\text{M}$  was present in the pipette solution (**e**, **f**).





**Figure 2** Intracellular 8-NH<sub>2</sub>-cADPR and heparin inhibit Ca<sup>2+</sup> release evoked by NAADP. Traces represent recordings from single cells from the same batch of cells tested in parallel. 8-NH<sub>2</sub>-cADPR (**b**) or heparin (**c**) was present in the pipette/intracellular solution at 18 μM and 250 μg ml<sup>-1</sup>, respectively. NAADP (**a**, **b**, **c**) was present in the pipette/intracellular solution at 50 nM.

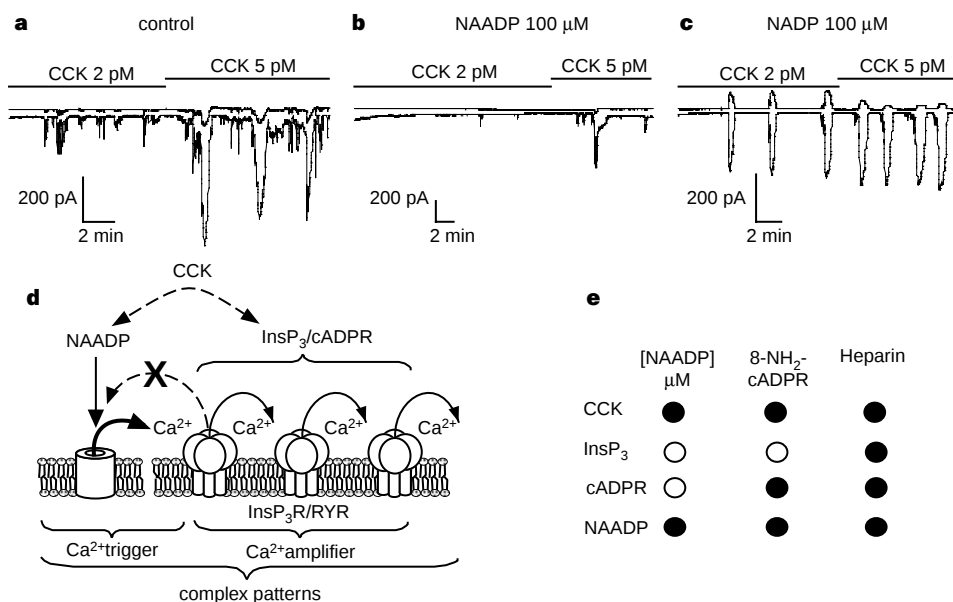
8-NH<sub>2</sub>-cADPR (18 μM) or heparin (250 μg ml<sup>-1</sup>) in the pipette solution abolished Ca<sup>2+</sup> spikes induced by NAADP (50 nM in the pipette) in pancreatic acinar cells (Fig. 2b, *n* = 5; Fig. 2c, *n* = 7, respectively). In contrast, in the absence of inhibitors, the cells responded vigorously to 50 nM NAADP (Fig. 2a, *n* = 9/10). These results indicate that the complex pattern of Ca<sup>2+</sup> spiking evoked by NAADP also involves activation of ryanodine and InsP<sub>3</sub> receptors. Because high concentrations of NAADP fully inhibited NAADP-evoked Ca<sup>2+</sup> spikes (Fig. 1f), we tested the specificity of this effect for NAADP-induced Ca<sup>2+</sup> release by investigating the ability of cADPR and InsP<sub>3</sub> to induce Ca<sup>2+</sup> spikes in the presence of an inactivating concentration of NAADP (100 μM). As described previously, in the presence of inositol 2,4,5-trisphosphate (Ins(2,4,5)P<sub>3</sub>; 10 μM in the patch-pipette solution), cells displayed typical short-lasting Ca<sup>2+</sup> spikes (1–2 s; Fig. 3a, *n* = 5)<sup>17,22,24</sup>. The pattern of spikes is similar to



**Figure 3** Effect of NAADP on Ca<sup>2+</sup> spikes evoked by Ins(2,4,5)P<sub>3</sub> and cADPR. Each pair of traces represents recordings from single or double cells from the same batch of cells tested in parallel. Ins(2,4,5)P<sub>3</sub> and cADPR were present in the pipette/intracellular solution at 10 μM. NAADP (**b**, **d**) was present in the pipette/intracellular solution at 100 μM.

those evoked by 10 μM cADPR (Fig. 3c, *n* = 5). When NAADP at 100 μM was present in the internal pipette solution, all the cells still responded to both Ins(2,4,5)P<sub>3</sub> (Fig. 3b, *n* = 5) and cADPR (Fig. 3d, *n* = 5). Taken together, these data indicate that NAADP's inhibition of NAADP-induced Ca<sup>2+</sup> spiking is probably due to a desensitization of the NAADP receptor rather than to inhibition of InsP<sub>3</sub> or ryanodine receptors.

In pancreatic acinar cells, agonist-induced Ca<sup>2+</sup> signals display patterns that are characteristic of the agonist producing them<sup>3,16,17,25</sup>. Recently, both InsP<sub>3</sub> and cADPR have been suggested to transduce CCK-receptor signalling, possibly through distinct receptor subtypes<sup>3,24</sup>. Nevertheless, NAADP-evoked Ca<sup>2+</sup> spikes can be intriguingly similar to those evoked by CCK at 2 pM and 5 pM (Fig. 4a). Taking advantage of the observation that high concentrations of NAADP selectively inhibit NAADP-induced Ca<sup>2+</sup> release, we tested



**Figure 4** Intracellular NAADP inhibits Ca<sup>2+</sup> spikes (short-lasting and long-lasting) evoked by physiological concentrations of CCK (2 and 5 pM). Traces represent recordings from single or double cells from the same batch of cells tested in parallel. CCK was applied outside the cells using a local perfusion system at the indicated concentration (**a**, **b**, **c**). NAADP (**b**) or NADP (**c**) was present in the pipette/intracellular solution at 100 μM in the indicated experiments. **d**, Model of NAADP-evoked Ca<sup>2+</sup> spiking. NAADP first releases Ca<sup>2+</sup> from the NAADP-sensitive store and then Ca<sup>2+</sup> recruits InsP<sub>3</sub> and ryanodine receptors (InsP<sub>3</sub>Rs and

RyRs) through Ca<sup>2+</sup>-induced Ca<sup>2+</sup> release in the presence of their respective agonists, InsP<sub>3</sub> and cADPR, triggering a complex pattern of Ca<sup>2+</sup> spikes. In contrast, direct activation of InsP<sub>3</sub> and ryanodine receptors by InsP<sub>3</sub> or cADPR results in short-lasting Ca<sup>2+</sup> spikes or a sustained Ca<sup>2+</sup> response<sup>3,17</sup>, and does not involve the NAADP-sensitive store. **e**, CCK and NAADP share a similar pharmacology; filled circles represent an inhibition of Ca<sup>2+</sup> release by antagonists and open circles no such inhibition.

whether NAADP was important in transducing  $\text{Ca}^{2+}$  signals mediated by CCK receptors. Extracellular application of CCK at 2 and 5 pM elicited typical  $\text{Ca}^{2+}$  spikes (Fig. 4a,  $n = 13/14$  and  $14/14$ , respectively). However, with NAADP (1–10  $\mu\text{M}$ ) in the intracellular pipette solution, only 57% of cells tested responded to 2 pM CCK, but all responded to 5 pM CCK ( $n = 7$ , not shown). A further increase in NAADP concentration to 100  $\mu\text{M}$  in the pipette solution completely abolished the ability of CCK (2 pM) to evoke a  $\text{Ca}^{2+}$  response (Fig. 4b,  $n = 7$ ). In addition, when we increased the CCK concentration to 5 pM, 57% of the cells did not respond and 43% displayed an attenuated response, indicating that the inhibitory effect of NAADP remains (Fig. 4b,  $n = 7$ ). This effect of NAADP was highly specific, as infusion of cells with its metabolic precursor  $\beta\text{-NADP}$  (100  $\mu\text{M}$ )<sup>11,26</sup> failed to modulate or inhibit the effects of CCK at either 2 or 5 pM, with  $\text{Ca}^{2+}$  spikes observed in all cells investigated (Fig. 4c,  $n = 4$ ). This indicates that NAADP synthesis from  $\beta\text{-NADP}$  must be tightly controlled. We found evidence that pancreatic acinar cells can synthesize NAADP from  $\beta\text{-NADP}$  by incubating saponin-permeabilized isolated cells with  $\beta\text{-NADP}$  in the presence of nicotinic acid<sup>15,26,27</sup>. We detected a NAADP-producing activity of  $2.55 \pm 0.08$  nmol per mg per h ( $n = 6$ ), consistent with the hypothesis that NAADP acts as an intracellular messenger in pancreatic acinar cells.

In this study, we report the first description of NAADP-evoked  $\text{Ca}^{2+}$  mobilization in a mammalian cell. The remarkably low concentrations of NAADP required to induce  $\text{Ca}^{2+}$  spiking and NAADP's ability to produce complex patterns of  $\text{Ca}^{2+}$  signals, together with the abolition of CCK signalling when cells are perfused with desensitizing concentrations of NAADP, indicates that NAADP may be a key intracellular mediator of this hormone. The inhibition of NAADP-induced spiking by  $\text{InsP}_3$  and cADPR antagonists suggests that a local NAADP-induced  $\text{Ca}^{2+}$  release through a non-regenerative mechanism<sup>14,28</sup> may then initiate  $\text{Ca}^{2+}$ -induced  $\text{Ca}^{2+}$  release through  $\text{InsP}_3$  and ryanodine receptors, giving rise to more global  $\text{Ca}^{2+}$  signals<sup>10</sup> (Fig. 4d, e). Because NAADP mobilizes  $\text{Ca}^{2+}$  in pancreatic acinar cells as well as in the oocytes of two diverse species of marine invertebrates<sup>7</sup>, it is probably a general  $\text{Ca}^{2+}$ -mobilizing messenger. As NAADP may be formed concomitantly with cADPR by ADP-ribosyl cyclases<sup>26,27</sup>, parallels can be drawn between the ADP-ribosyl cyclase<sup>7</sup> and phospholipase C signalling pathways<sup>1</sup> in producing multiple messengers that regulate diverse cellular functions by generating specific  $\text{Ca}^{2+}$  signalling patterns. □

## Methods

**Pancreatic acinar cells.** Isolated single and double mouse pancreatic acinar cells were prepared as previously described<sup>16,17</sup>.

**Patch-clamp recordings.** Cells were investigated using the whole-cell patch-clamp configuration. From a holding potential of  $-30$  mV, steps were made to 0 mV, the reversal potential of the two  $\text{Ca}^{2+}$ -dependent currents<sup>16,17,22,27,24</sup>. Using our solutions, the reversal potential of both the  $\text{Cl}^-$  and non-selective cation currents was 0 mV (for use as a potential control). Small deviations in chloride and cation potentials and in the holding potential sometimes produce small inward or outward currents at 0 mV. At  $-30$  mV we obtained a measure of both the  $\text{Ca}^{2+}$  currents, which are an index of cytosolic  $\text{Ca}^{2+}$  changes<sup>16,17</sup>. The extracellular  $\text{Na}^+$ -rich solution contained, in mM: 140 NaCl, 4.7 KCl, 1.13  $\text{MgCl}_2$ , 10 glucose, 1  $\text{CaCl}_2$  and 10 HEPES–NaOH, pH 7.2. CCK octapeptide was added to the external solution as indicated. The  $\text{Ca}^{2+}$ -free solution was obtained either by removing external  $\text{Ca}^{2+}$  or by using an external solution containing 10 mM EGTA. The internal solution contained, in mM: 140 KCl, 1.13  $\text{MgCl}_2$ , 0.05 EGTA, 2 ATP and 10 HEPES–KOH, pH 7.2. The  $\text{Ca}^{2+}$  concentration in the patch-pipette solution was measured using the  $\text{Ca}^{2+}$ -sensitive dye fura-2 and the ratiometric equation with a dissociation constant of 135 nM<sup>29</sup>. The  $\text{Ca}^{2+}$  concentration in the internal solution containing 50  $\mu\text{M}$  EGTA and 50 nM NAADP was 130 nM. Caged NAADP, NAADP, cADPR and 8- $\text{NH}_2$ -cADPR were from Molecular Probes. All other chemicals and NAADP were from Sigma.

**$\text{Ca}^{2+}$  release assays.** Homogenates of sea urchin eggs were prepared and used as a bioassay as described previously<sup>6</sup>.

**Detection of NAADP metabolism.** The ability of pancreatic acinar cells to synthesize NAADP was determined by incubating permeabilized pancreatic acinar cells (saponin, 50  $\mu\text{g ml}^{-1}$ ) in the presence of  $\beta\text{-NADP}^+$  (250  $\mu\text{M}$ ) and nicotinic acid (14 mM). The synthesis of NAADP was monitored over time by removing 5  $\mu\text{l}$  of the reaction and testing it using the bioassay system<sup>27</sup>.

**Flash photolysis.** The caged NAADP at 100  $\mu\text{M}$  was photolysed with a 1-ms ultraviolet flash (250 kW) from a xenon flash lamp (XF-10, HI-TECH Scientific) passed through a 300–400-nm band-pass filter. The quantity of the photolysed product was compared with a known concentration of authentic NAADP and its identity confirmed using cross-desensitization with standard NAADP in sea urchin homogenate. The ultraviolet flash efficiency was 0.5–1%.

Received 19 November 1998; accepted 19 January 1999.

- Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–325 (1993).
- Cobbold, P. H., Sanchez-Bueno, A. & Dixon, C. J. The hepatocyte calcium oscillator. *Cell Calcium* **12**, 87–95 (1991).
- Petersen, O. H., Petersen, C. C. H. & Kasai, H. Calcium and hormone action. *Annu. Rev. Physiol.* **56**, 297–319 (1994).
- Thomas, A. P., Bird, G. S. J., Hajnoczky, G., Robb-Gasper, L. d. & Putney, J. W. Spatial and temporal aspects of cellular calcium signaling. *FASEB J.* **10**, 1505–1517 (1996).
- Tsien, R. W. & Tsien, R. Y. Calcium channels, stores, and oscillations. *Annu. Rev. Cell Biol.* **6**, 715–760 (1990).
- Gallione, A., Lee, H. C. & Busa, W. B.  $\text{Ca}^{2+}$ -induced  $\text{Ca}^{2+}$  release in sea urchin egg homogenates and its modulation by cyclic ADP-ribose. *Science* **253**, 1143–1146 (1991).
- Lee, H. C. Mechanisms of calcium signalling by cyclic ADP-ribose and NAADP. *Physiol. Rev.* **77**, 1133–1164 (1997).
- Taylor, C. & Marshall, I. Calcium and inositol 1,4,5-trisphosphate receptors: a complex relationship. *Trends Biochem. Sci.* **17**, 403–407 (1992).
- Meissner, G. Ryanodine receptor/ $\text{Ca}^{2+}$  release channels and their regulation by endogenous effectors. *Annu. Rev. Physiol.* **56**, 485–508 (1994).
- Berridge, M. J. Elementary and global aspects of calcium signalling. *J. Physiol. (Lond.)* **499**, 291–306 (1997).
- Lee, H. C. & Aarhus, R. A derivative of NADP mobilizes calcium stores insensitive to inositol trisphosphate and cyclic ADP-ribose. *J. Biol. Chem.* **270**, 2152–2157 (1995).
- Chini, E. N., Beers, K. W. & Dousa, T. P. Nicotinic adenine dinucleotide phosphate (NAADP) triggers a specific calcium release system in sea urchin eggs. *J. Biol. Chem.* **270**, 3216–3223 (1995).
- Albrieux, M., Lee, H. C. & Villaz, M. Calcium signaling by cyclic ADP-ribose, NAADP, and inositol trisphosphate are involved in distinct functions in ascidian oocytes. *J. Biol. Chem.* **273**, 14566–14574 (1998).
- Genazzani, A. A. & Gallione, A. Nicotinic acid adenine dinucleotide phosphate releases  $\text{Ca}^{2+}$  from a thapsigargin-insensitive pool. *Biochem. J.* **315**, 721–725 (1996).
- Genazzani, A. A. & Gallione, A. A  $\text{Ca}^{2+}$  release mechanism gated by the novel pyridine nucleotide, NAADP. *Trends Pharmacol. Sci.* **18**, 108–110 (1997).
- Ospichuk, Y. V., Wakui, M., Yule, D. I., Gallacher, D. V. & Petersen, O. H. Cytoplasmic  $\text{Ca}^{2+}$  oscillations evoked by receptor stimulation, G-protein activation, internal application of inositol trisphosphate or  $\text{Ca}^{2+}$ : simultaneous microfluorimetry and  $\text{Ca}^{2+}$  dependent  $\text{Cl}^-$  current recording in single pancreatic acinar cells. *EMBO J.* **9**, 697–704 (1990).
- Thorn, P., Lawrie, A. M., Smith, P. M., Gallacher, D. V. & Petersen, O. H. Local and global cytosolic  $\text{Ca}^{2+}$  oscillations in exocrine cells evoked by agonist and inositol trisphosphate. *Cell* **74**, 661–688 (1993).
- Genazzani, A. A., Empson, R. M. & Gallione, A. Unique inactivation properties of NAADP-sensitive  $\text{Ca}^{2+}$  release. *J. Biol. Chem.* **271**, 11599–11602 (1996).
- Aarhus, R. et al. Activation and inactivation of  $\text{Ca}^{2+}$  release by NAADP. *J. Biol. Chem.* **271**, 8513–8516 (1996).
- Lee, H. C., Aarhus, R., Gee, K. R. & Kestner, T. Caged nicotinic acid adenine dinucleotide phosphate. *J. Biol. Chem.* **272**, 4172–4178 (1997).
- Walseth, T. F. & Lee, H. C. Synthesis and characterization of antagonists of cyclic ADP-ribose. *Biochim. Biophys. Acta* **1178**, 235–242 (1993).
- Cancela, J. M. & Petersen, O. H. The cyclic ADP-ribose antagonist 8- $\text{NH}_2$ -cADP-ribose blocks cholecystokinin-evoked cytosolic  $\text{Ca}^{2+}$  spiking in pancreatic acinar cells. *Pflügers Arch.* **435**, 746–748 (1998).
- Ehrlich, B. E., Kaftan, E., Bezprozvannaya, S. & Bezprozvanny, I. The pharmacology of intracellular  $\text{Ca}^{2+}$  release channels. *Trends Pharmacol. Sci.* **15**, 145–148 (1994).
- Cancela, J. M., Mogami, H., Tepikin, A. V. & Petersen, O. H. Intracellular glucose switches between cyclic ADP-ribose and inositol trisphosphate triggering of cytosolic  $\text{Ca}^{2+}$  spiking. *Curr. Biol.* **8**, 865–868 (1998).
- Petersen, C. C. H., Toescu, E. C. & Petersen, O. H. Different patterns of receptor-activated cytoplasmic  $\text{Ca}^{2+}$  oscillations in single pancreatic acinar cells: dependence on receptor type, agonist concentration and intracellular  $\text{Ca}^{2+}$  buffering. *EMBO J.* **10**, 527–533 (1991).
- Aarhus, R., Graeff, R. M., Dickey, D. M., Walseth, T. F. & Lee, H. C. ADP-ribosyl cyclase and CD38 catalyze the synthesis of a calcium-mobilizing metabolite from NADP<sup>+</sup>. *J. Biol. Chem.* **270**, 30327–30333 (1995).
- Wilson, H. L. & Gallione, A. Differential regulation of nicotinic acid-adenine dinucleotide phosphate and cADP-ribose production by cAMP and cGMP. *Biochem. J.* **331**, 837–843 (1998).
- Chini, E. N. & Dousa, T. P. Nicotinate-adenine dinucleotide phosphate-induced  $\text{Ca}^{2+}$  release does not behave as a  $\text{Ca}^{2+}$ -induced  $\text{Ca}^{2+}$ -release system. *Biochem. J.* **316**, 709–711 (1996).
- Gryniewicz, G., Poenie, M. & Tsien, R. Y. A new generation of calcium indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260**, 3440–3450 (1985).

**Acknowledgements.** We thank A. David Smith for comments on the manuscript. This work was supported by the EU (J.M.C., A.G.) and the Wellcome Trust (G.C.C., A.G.).

Correspondence should be addressed to A.G. (e-mail: antony.gallione@pharm.ox.ac.uk) or J.M.C. (j.m.cancela@liverpool.ac.uk). Requests for materials should be addressed to G.C.C. (e-mail: grant.churchill@pharm.ox.ac.uk).

# Cytotoxic T-cell immunity to virus-infected non-haematopoietic cells requires presentation of exogenous antigen

Luis J. Sigal\*, Shane Crotty†, Raul Andino† & Kenneth L. Rock\*

\* Department of Pathology, University of Massachusetts Medical Center, Worcester, Massachusetts 01655-0118, USA

† Department of Microbiology and Immunology, University of California, Box 0414, San Francisco, California 94143-0414, USA

Cytotoxic T lymphocytes (CTLs) are thought to detect viral infections by monitoring the surface of all cells for the presence of viral peptides bound to major histocompatibility complex (MHC) class I molecules. In most cells, peptides presented by MHC class I molecules are derived exclusively from proteins synthesized by the antigen-bearing cells<sup>1</sup>. Macrophages and dendritic cells also have an alternative MHC class I pathway that can present peptides derived from extracellular antigens; however, the physiological role of this process is unclear<sup>2</sup>. Here we show that virally infected non-haematopoietic cells are unable to stimulate primary CTL-mediated immunity directly. Instead, bone-marrow-derived cells are required as antigen-presenting cells (APCs) to initiate anti-viral CTL responses. In these APCs, the alternative (exogenous) MHC class I pathway is the obligatory mechanism for the initiation of CTL responses to viruses that infect only non-haematopoietic cells.

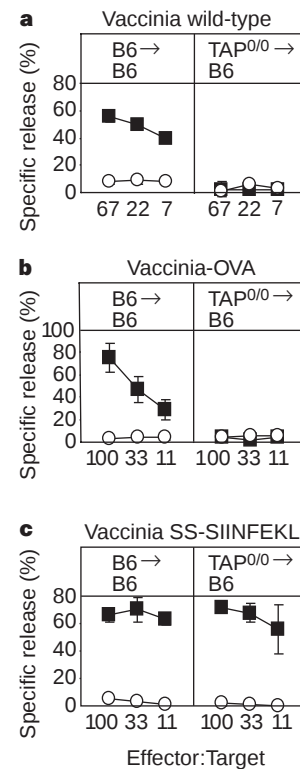
The 'classical' MHC class I antigen-presentation pathway is thought to be the major mechanism used by the immune system to detect viral infections in all cells. In this pathway, proteins synthesized by a cell are degraded in the cytoplasm into oligopeptides, a fraction of which are transported into the endoplasmic reticulum (ER) by the transporter associated with antigen presentation (TAP). In the ER these peptides bind to new MHC class I molecules and the resulting complexes are transported to the cell surface. As MHC class I molecules must bind peptides in order to be transported to the plasma membrane, TAP is required for normal MHC class I expression at the cell surface and for antigen presentation<sup>1</sup>.

To determine whether non-haematopoietic cells can function as APCs to initiate CTL responses to viruses, we constructed bone-marrow chimaeras by lethally irradiating C57Bl/6 mice (B6 mice; MHC class I haplotype H-2<sup>b</sup>) and reconstituting them with bone marrow from TAP<sup>0/0</sup> mice<sup>3</sup> (also H-2<sup>b</sup>; all bone-marrow chimaeras will be referred to as bone-marrow donor → irradiated recipient). As bone-marrow-derived cells in TAP<sup>0/0</sup> → B6 mice cannot transfer peptides from the cytosol to the ER, they are unable to use the classical MHC class I pathway<sup>3,4</sup>.

The chimaeric mice were assayed for the generation of CTL responses following infection with wild-type vaccinia virus or with vaccinia-OVA, a recombinant vaccinia virus carrying chicken ovalbumin (OVA) as a full-length protein<sup>5</sup>. Although they have intact MHC class I antigen presentation in non-haematopoietic tissues, TAP<sup>0/0</sup> → B6 mice did not generate CTL responses to vaccinia-viral antigens (Fig. 1a) or to OVA (Fig. 1b). In contrast, robust CTL responses to these antigens were detected in control B6 → B6 mice. These results indicate that the generation of CTL responses to vaccinia virus requires bone-marrow-derived cells with functional TAP molecules.

Next, we determined whether the inability of TAP<sup>0/0</sup> → B6 mice to generate CTL responses was due to a defect in CD8<sup>+</sup> T cells or to a

failure in antigen presentation by bone-marrow-derived APCs. TAP<sup>0/0</sup> (non-chimaeric) mice almost completely lack CD8<sup>+</sup> T cells because expression of MHC class I molecules in epithelial thymic cells is necessary for the positive selection of CD8<sup>+</sup> T cells in the thymus<sup>3</sup>. However, as shown by flow-cytometry analysis, TAP<sup>0/0</sup> → B6 mice had normal numbers of CD8<sup>+</sup> T cells in peripheral blood and spleen (ref. 4 and data not shown) as a consequence of positive selection on wild-type thymic epithelial cells (which are radioresistant and non-haematopoietic). These CD8<sup>+</sup> T cells were fully functional. TAP<sup>0/0</sup> → B6 mice and control B6 → B6 mice generated a similar CTL response to vaccinia-SS-SIINFEKL<sup>5</sup> (Fig. 1c), a recombinant vaccinia virus expressing the antigenic peptide of OVA (SIINFEKL)<sup>6</sup> preceded by a signal sequence that delivers the peptide directly into the ER, thus bypassing the need for TAP. TAP<sup>0/0</sup> → B6 chimaeras and control mice also generated anti-SIINFEKL CTL when injected intravenously with a graded number of B6 dendritic cells that had been preincubated with a single concentration of synthetic SIINFEKL (not shown) or when injected with 5 × 10<sup>5</sup> dendritic cells that had been incubated with graded concentrations of SIINFEKL (Fig. 2). In these latter experiments, cells incubated with 10 nM SIINFEKL and cells infected *in vitro* with vaccinia-OVA and polio-OVA (see below) stimulated a T-cell



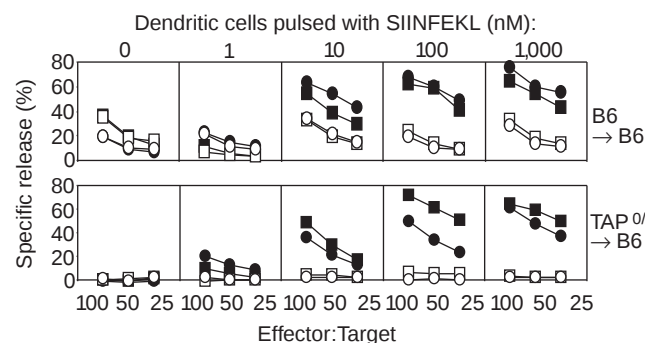
**Figure 1** CTL responses to wild-type vaccinia virus and OVA in recombinant vaccinia virus requires a bone-marrow-derived APC. **a**, CTL response to wild-type vaccinia. The indicated bone-marrow chimaeras were infected with  $2 \times 10^7$  p.f.u. of wild-type vaccinia virus. One week later, mice were killed and freshly explanted spleen cells were assayed in <sup>51</sup>Cr-release assays on vaccinia-infected MC57G cells (filled squares) or uninfected MC57G cells as controls (open circles). The TAP<sup>0/0</sup> → B6 mice did not generate a response to the vaccinia-viral antigens. **b**, CTL response to OVA in vaccinia-OVA protein. The indicated bone-marrow chimaeras were infected with  $2 \times 10^7$  p.f.u. of vaccinia-OVA. One week later, mice were killed and their spleen cells were cultured in the presence of mitomycin-C (Sigma)-treated EG7 cells<sup>29</sup> (an EL-4-derived cell line stably transfected with OVA). After five days cells were collected and tested in <sup>51</sup>Cr-release assays on EG7 targets (filled squares) on EL-4 targets as controls (open circles). **c**, CTL response to vaccinia-SS-SIINFEKL. The experiment was performed as in **b** except that mice were infected with vaccinia-SS-SIINFEKL. The x-axis shows the ratio of effector cells to target cells.



hybrid specific for the peptide-MHC complex SIINFEKL-K<sup>b</sup> at roughly similar levels (data not shown).

These results indicate that the inability of TAP<sup>0/0</sup> → B6 mice to generate a CTL response to vaccinia virus is due to a failure of antigen presentation. This failure to present antigens did not result from an inability of the TAP<sup>0/0</sup> → B6 chimaeras to reconstitute bone-marrow-derived APCs, because these mice generated MHC class II-restricted T-cell responses to OVA that were at least as strong as those generated by B6 → B6 animals (Fig. 3a, b). Together these data show that bone-marrow-derived professional APCs, possessing a functional TAP, are required to initiate CTL responses to vaccinia, and that non-haematopoietic tissues infected with vaccinia cannot prime CTLs, despite the ability of vaccinia to infect many different tissues, including respiratory organs, liver, kidney, spleen, ovaries and the central nervous system<sup>7-9</sup>. There may be several explanations for the inability of non-haematopoietic cells to stimulate a CTL response. Although non-haematopoietic cells are able to present antigenic peptides bound to MHC class I, they express low levels of these molecules in the absence of inflammation. Moreover, they do not express MHC class II molecules, which are essential for the stimulation of CD4<sup>+</sup> helper T cells, and they lack adhesion and co-stimulatory molecules that might be required to stimulate naive T cells. Non-immune cells also lack the ability to migrate to lymphoid organs, where many immune responses are initiated<sup>10</sup>. On the other hand, some bone-marrow-derived cells, such as macrophages and dendritic cells, express high levels of MHC class I and class II molecules and several adhesion and co-stimulatory molecules, including B7.1 and B7.2, and can migrate to central lymphoid organs. However, after naive CTLs are stimulated to become effectors, they no longer require co-stimulation or T-cell help and can recognize lower levels of peptide-MHC complexes. Therefore, once stimulated by professional APCs, the effector CTLs acquire the ability to migrate out of the lymphoid organs to clear viral infections in all tissues.

In the experimental model described above, bone-marrow-derived APCs might acquire the viral antigens by becoming infected themselves<sup>11</sup>. In fact, SIINFEKL was presented by vaccinia-OVA infected dendritic cells and macrophages *in vitro* (not shown). However, it seems unlikely that professional APCs would become infected by all viruses and therefore this mechanism would be unavailable to detect infections by many tissue-specific viruses. Alternatively, the bone-marrow-derived APCs might acquire vaccinia-viral antigens exogenously from other antigen-bearing

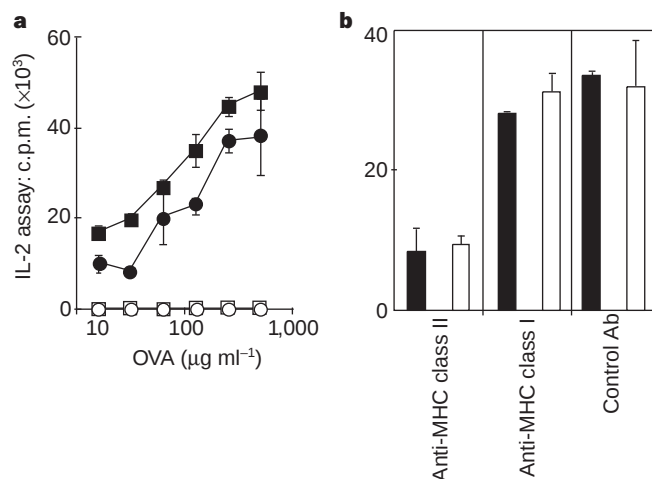


**Figure 2** TAP<sup>0/0</sup> → B6 mice can generate CTL responses comparable to those in B6 → B6 mice when immunized with APCs loaded with antigenic peptide. The indicated chimaeric mice were immunized intravenously with  $5 \times 10^5$  B6-derived, *in vitro*-cultured dendritic cells that had been incubated with SIINFEKL peptide at the indicated concentrations. One week later mice were killed and their spleen cells restimulated *in vitro* for 4 days with mitomycin-C-treated EL-4 cells that had been incubated with SIINFEKL. For each panel, squares represent one mouse and circles represent another. Targets were EL-4 cells that had been preincubated with SIINFEKL (filled symbols) or EL-4 cells that had not been incubated with peptide as controls (open symbols).

cells and this mechanism could operate in all infections, as proposed previously<sup>2,12</sup>. To determine whether the presentation of exogenous antigen is important in viral immunity, we developed a model in which bone-marrow-derived cells could not be infected with a virus.

Poliovirus (polio) is a positive-strand RNA virus with a host range that includes humans but not mice. This host-range restriction is determined by the expression of a suitable poliovirus receptor (PVR) on host cells<sup>13</sup>. Cells from wild-type mice can not be infected with polio, but mouse cells transfected with human PVR can be infected (ref. 14 and data not shown). In this study we used two human PVR-transgenic mice. We constructed a transgenic mouse (cPVR, in an ICR background) expressing a full-length PVR complementary DNA, driven by the  $\beta$ -actin promoter, in all tissues studied. Such mice are susceptible to polio infection and die with poliomyelitis following injection with wild-type polio (not shown). Following intraperitoneal infection of cPVR and control (B6) mice with the wild-type strain of polio, we found much higher titres of virus (expressed as plaque-forming units per tissue) in the skeletal muscle (12,000), brain (1,600) and spinal cord (19,000), and slightly higher titres in the kidney (1.3) and liver (0.64), of the cPVR mice (B6-mouse titres: 0.70, <0.60, <0.30, <0.10 and <0.20, respectively). We mated cPVR mice with B6 mice, and their progeny, (cPVR × B6) F<sub>1</sub> mice (referred to here as cPVR mice), were used here. Another transgenic mouse strain (referred to here as gPVR) possesses the human PVR genomic locus, including the endogenous human promoter, backcrossed onto the B6 background. It also supports viral replication in skeletal muscle and central nervous system<sup>13,15</sup>.

To examine the role of the exogenous MHC class I pathway in the initiation of CTL responses, we generated a series of bone-marrow-chimaeric mice, including two sets (B6 → cPVR and B6 → gPVR) that allowed infection of only non-bone-marrow-derived cells. In a previous study<sup>14</sup>, we constructed a polio virus recombinant (polio-OVA) expressing the carboxy-terminal half of OVA (which includes the SIINFEKL epitope). In this construct, the OVA fragment is synthesized as part of the viral polypeptide and is released in the cytosol by viral proteinases. We showed that polio-OVA can induce anti-OVA CTL responses in gPVR mice and cPVR but not in B6 mice (ref. 14 and data not shown). Consistent with this result,



**Figure 3** TAP<sup>0/0</sup> → B6 mice can generate MHC class II-restricted responses comparable to those in B6 → B6 mice. **a**, Production of IL-2 by lymph-node cells of immunized chimaeras in response to different concentrations of OVA. Cells were from OVA-immunized TAP<sup>0/0</sup> → B6 mice (filled squares) and B6 → B6 mice (filled circles) or from unimmunized controls (open squares and open circles respectively). **b**, The same cells as those used in **a** were incubated with 0.5 mg ml<sup>-1</sup> OVA in the presence of the indicated antibody-containing supernatants. Only the results for immunized TAP<sup>0/0</sup> → B6 mice (filled columns) or B6 → B6 mice (open columns) are shown.

cPVR → cPVR mice infected with polio-OVA generated anti-OVA CTLs (Fig. 4Ac), but B6 → B6 mice did not (Fig. 4Ad). B6 → cPVR chimaeric mice, which have bone-marrow-derived cells that cannot be infected by polio (PVR-negative cells), generated strong anti-OVA CTL responses when infected with polio-OVA (Fig. 4Aa) but not when infected with a recombinant polio (polio-sp27) expressing an irrelevant protein (Fig. 4Ae). Identical results were obtained when using gPVR recipients (Fig. 4Ba, c and e). Therefore, either the infected non-haematopoietic cells are stimulating CTL responses, or bone-marrow-derived cells are acquiring the polio-expressed OVA from exogenous sources.

To distinguish between these two possibilities, we constructed chimaeric mice by using TAP<sup>0/0</sup> mice as bone-marrow donors and PVR<sup>+</sup> transgenic mice as recipients. Remarkably, TAP<sup>0/0</sup> → cPVR (Fig. 4Ab) and TAP<sup>0/0</sup> → gPVR (Fig. 4Bb) mice did not generate anti-OVA CTL responses when infected with polio-OVA. As expected, their CTLs were functional and generated a strong response to vaccinia-SS-SIINFEKL (data not shown). In contrast to the control B6 → gPVR mice, these TAP<sup>0/0</sup> chimaeric mice also

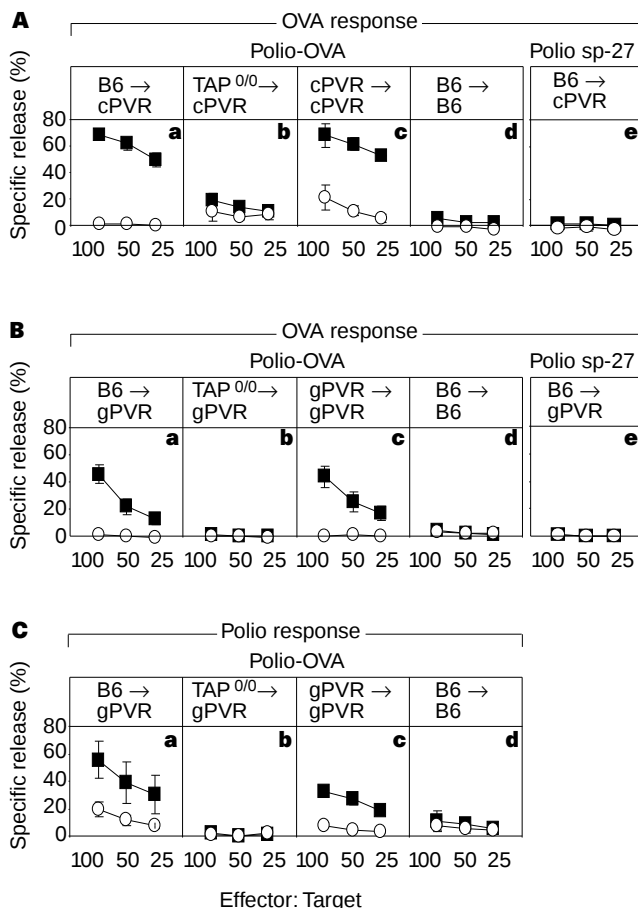
failed to generate CTL responses to a D<sup>b</sup>-presented epitope from the polio VP0 protein (Fig. 4C). As shown above with vaccinia virus, these data indicate that non-haematopoietic cells are unable to stimulate CTL immunity to another virus, indicating that this may be a general rule. These results also indicate that the bone-marrow-derived cells in B6 → c/gPVR mice acquired antigen from exogenous sources. That TAP<sup>0/0</sup> → c/gPVR mice did not respond to either antigen also indicates that the response in B6 → c/gPVR mice was not due to residual PVR<sup>+</sup> bone-marrow-derived cells that survived irradiation.

Until now, the physiological function of the exogenous MHC class I pathway has been unclear. Stimulation of CTLs by this route has been shown to occur in several situations, such as transplantation ('crosspriming' for minor histocompatibility antigens)<sup>16</sup> and injection of particulate antigens<sup>17</sup>, but it has been thought to make a minor contribution to overall responses. Two situations in which this pathway has been shown to be important are in the generation of CTL responses to a tumour<sup>4</sup> and in the homing and development of tolerance of adoptively transferred T cells specific for a transgenic antigen expressed in pancreatic  $\beta$ -cells<sup>18,19</sup>. However, in these cases it was unclear whether the exogenous pathway might be dominant only because of the lack of inflammation (which stimulates antigen presentation and provides an adjuvant effect) and/or because these cells might be poor stimulators. It has been suggested that the exogenous MHC class I pathway is inefficient and unlikely to play an important part in most physiological situations<sup>20</sup>. Our experiments with B6 → c/gPVR mice contradict this view and indicate that the exogenous MHC class I pathway is essential for the initiation of CTL responses to viral infection that is confined to non-haematopoietic tissues. In fact, if this pathway did not exist, viruses could escape immune surveillance by using receptors that are not expressed on the critical, bone-marrow-derived APCs. Our results indicate that the presentation of exogenous antigen is a major pathway *in vivo* and may contribute to the stimulation of CTL responses even in situations in which viruses may infect bone-marrow-derived cells.

How do bone-marrow-derived cells acquire viral exogenous antigens? When infected cells die *in vivo*, they are rapidly cleared by bone-marrow-derived phagocytes, which will import the viral antigens into the exogenous MHC class I and class II pathways<sup>21</sup>. Interestingly, antigens from apoptotic cells are avidly presented on class I molecules of dendritic cells<sup>22</sup>. Although our results do not specifically establish the identity of the bone-marrow-derived APCs responsible for initiating CTL responses through the exogenous pathway, macrophages and/or dendritic cells are again the likely candidates, because they can present antigen through the exogenous MHC class I pathway *in vitro*<sup>17,23</sup>. These cells can also ingest dying cells and cellular debris by phagocytosis and can thereby import viral antigens into the exogenous MHC class I pathway. Moreover, their migratory nature allows them to acquire antigen at a site of infection and then travel to the lymphoid tissues.

Two routes for the exogenous MHC class I pathway *in vitro* have been described, a TAP-independent pathway, in which antigen is probably hydrolysed in endosomes<sup>24,25</sup>, and a phagosome-to-cytosol pathway<sup>26</sup> that is TAP-dependent. Our data provide indirect evidence that, *in vivo*, vaccinia-OVA and polio-OVA antigens may follow the TAP-dependent exogenous MHC class I pathway.

Our results show a strict requirement for professional APCs in the generation of anti-viral CTL immunity, and demonstrate that the exogenous pathway plays a key part in the immune surveillance of non-haematopoietic tissues. These findings have implications for vaccine delivery and gene therapy as well as for immune evasion by viruses. The results indicate that, to stimulate strong immunity, viral vectors or naked DNA must be expressed in professional APCs or delivered in a manner that will promote exogenous antigen presentation. Moreover, these mechanisms may limit the ability of viruses to block the generation of CTLs by downregulating MHC class I expression on infected cells<sup>27</sup>, because this is unlikely to affect the exogenous pathway in uninfected professional APCs. □



**Figure 4** Initiation of the CTL response to polio-OVA requires the presence of bone-marrow-derived APCs, but not their infection. Each panel corresponds to a single experiment. **A, B**, Bone-marrow chimaeras (**Aa–e**, cPVR recipients and B6 control recipients; **Ba–e**, gPVR recipients and B6 control recipients) were infected with polio-OVA or polio-sp27 as indicated. Mice were killed 3 weeks later and their spleen cells were co-cultured for 5 days with mitomycin-C-treated EG7 cells and used in <sup>51</sup>Cr-release assays. Filled squares, EG7 targets; open circles, EL-4 targets as controls. **Ca–d**, Bone-marrow chimaeras obtained using gPVR recipients and B6 control recipients were infected with polio-OVA. Three weeks later, mice were killed and their spleen cells were cultured in the presence of mitomycin-C-treated EL-4 cells that had been preincubated with 2  $\mu$ g ml<sup>-1</sup> of the D<sup>b</sup>-binding polio-derived peptide P22. After 5 days, cells were collected and tested in <sup>51</sup>Cr-release assays. Targets were EL-4 cells preincubated with P22 (filled squares) or EL-4 cells without peptide as controls (open circles).

## Methods

TAP<sup>0/0</sup> (B6,129-Tap<sup>1AP</sup>; Jackson Laboratory) and B6 (Taconic) mice were obtained at 6–8 weeks of age. cPVR mice were made by standard transgenic techniques using the plasmid pVR-9, which has already been described<sup>14</sup>. gPVR mice (a gift from Cynamid) were bred at UMMC animal facilities. To prepare chimaeras, bone-marrow cells from 1–3-month-old donor mice were treated with anti-Thy1 antibody (M5/49.4.1; ATCC) and complement to eliminate mature T cells, washed twice and resuspended in PBS. Recipients were irradiated with 650 rad and then with 450 rad 4 h later. Irradiated mice were reconstituted by intravenous inoculation of 4–6 × 10<sup>6</sup> bone-marrow cells from the different donors. To avoid rejection of donor MHC class I-negative TAP<sup>0/0</sup> cells by host natural-killer cells, chimaeras also received an intraperitoneal injection of 10 µl rabbit anti-asialo GM1 gammaglobulin (Wako Chemicals) on the day of the transplant, and a second injection 3 days later. Bone-marrow chimaeras were rested for 4–6 months following reconstitution to allow for complete elimination of host-derived professional APCs. Mice were inoculated with virus and CTL killing was measured from fresh spleen cells (wild-type vaccinia), or from cultures restimulated with antigen for 5 days, using a <sup>51</sup>Cr-release assay as described<sup>28</sup>. For MHC class II-restricted responses, mice were injected at the base of the tail with 100 µg OVA (Sigma) emulsified in complete Freund's adjuvant (Gibco) in a final volume of 50 µl. Ten days later mice were killed and 2 × 10<sup>5</sup> cells from pooled para-aortic lymph nodes were incubated for 48 h in triplicate wells of microtitre plates in the presence of OVA. Supernatants were assayed for the presence of interleukin (IL)-2 by measuring the incorporation of <sup>3</sup>H-thymidine by the IL-2-dependent cell line CTLL. When indicated, antibodies were added to culture as a 1:8 dilution of hybridoma culture supernatants. The antibodies M5/114, Y-3 and BBM.1 (ATCC) were used as anti-MHC class II, anti-MHC class I and control antibodies, respectively. All experiments were performed at least three times. Data points in all figures except Fig. 2 represent averages ± s.e.m. for three mice (all experimental groups) or two mice (vaccinia-SS-SIINFEKL; polio-sp27 controls). In Fig. 2, two mice were used per group and results from these mice are shown individually. All mice used in these experiments were housed at the UMMC animal facilities and experiments were conducted in compliance with NIH and institutional guidelines.

Viruses were produced and used as described<sup>14,28</sup> except that inoculation of mice with polio was performed intravenously rather than intraperitoneally. For the generation of CTL responses to all viruses, the inoculation dose was 2 × 10<sup>7</sup> plaque-forming units (p.f.u.) per mouse diluted in 0.5 ml PBS. For the recovery of infectious virus from organs, mice were inoculated intraperitoneally with 2 × 10<sup>8</sup> p.f.u. of the poliovirus wild-type Mahoney 1 strain. Six paralysed cPVR mice were killed at days 4.5–6.5 after infection. Four control B6 mice were killed at days 4.5–5.5. Tissue samples were homogenized, and poliovirus titres in each tissue were determined by plaque assay.

P22 is a D<sup>b</sup>-binding peptide corresponding to amino acids 22–30 of the poliovirus polyprotein that we have recently identified (L.J.S. and K.L.R., manuscript in preparation). Both SIINFEKL and P22 were synthesized at the peptide core facility at UMMC.

All cell lines used in this study have been described and were cultured as previously<sup>28</sup>. To obtain dendritic cells, bone-marrow cells obtained from B6 mice were incubated overnight in RPMI media (Irvine Scientific) supplemented with 10% fetal calf serum (Atlanta Biologicals), 5 × 10<sup>-5</sup> M 2-mercaptoethanol (Sigma) and 2 mM L-glutamine, antibiotics (Fungi-Bact), 0.01 M HEPES buffer and non-essential amino acids (all from Irvine Scientific). Non-adherent cells were collected and grown in the same media supplemented with 10 ng ml<sup>-1</sup> granulocyte/macrophage colony-stimulating factor and 5 ng ml<sup>-1</sup> IL-4 (Pharmingen) for 5–6 days with further addition of cytokines every other day. Most cells in these cultures were dendritic cells, as judged by morphology and analysis expression of specific markers by flow cytometry. Before being injected into mice, 3 × 10<sup>6</sup> dendritic cells were thoroughly washed in PBS, resuspended in 1 ml PBS containing the indicated concentrations of SIINFEKL and 10 µg human β<sub>2</sub>-microglobulin (Calbiochem) and incubated at 37 °C for 1 h. Following incubation the cells were washed in PBS and injected intravenously into mice.

Received 23 July 1998; accepted 18 January 1999.

1. York, I. A. & Rock, K. L. Antigen processing and presentation by the class I major histocompatibility complex. *Annu. Rev. Immunol.* **14**, 369–396 (1996).

2. Rock, K. L. A new foreign policy: MHC class I molecules monitor the outside world. *Immunol. Today* **17**, 131–137 (1996).
3. Van Kaer, L., Ashton-Rickardt, P. G., Ploegh, H. L. & Tonegawa, S. TAP1 mutant mice are deficient in antigen presentation, surface class I molecules, and CD4-8+ T cells. *Cell* **71**, 1205–1214 (1992).
4. Huang, A. Y., Bruce, A. T., Pardoll, D. M. & Levitsky, H. I. *In vivo* cross-priming of MHC class I-restricted antigens requires the TAP transporter. *Immunity* **4**, 349–355 (1996).
5. Restifo, N. P. *et al.* Antigen processing *in vivo* and the elicitation of primary CTL responses. *J. Immunol.* **154**, 4414–4422 (1995).
6. Rotzschke, O. *et al.* Exact prediction of a natural T cell epitope. *Eur. J. Immunol.* **21**, 2891–2894 (1991).
7. Lee, M. S. *et al.* Molecular attenuation of vaccinia virus: mutant generation and animal characterization. *J. Virol.* **66**, 2617–2630 (1992).
8. Taylor, G., Stott, E. J., Wertz, G. & Ball, A. Comparison of the virulence of wild-type thymidine kinase (tk)-deficient and tk+ phenotypes of vaccinia virus recombinants after intranasal inoculation of mice. *J. Gen. Virol.* **72**, 125–130 (1991).
9. Buller, R. M., Smith, G. L., Cremer, K., Notkins, A. L. & Moss, B. Decreased virulence of recombinant vaccinia virus expression vectors is associated with a thymidine kinase-negative phenotype. *Nature* **317**, 813–815 (1985).
10. Kundig, T. M. *et al.* Fibroblasts as efficient antigen-presenting cells in lymphoid organs. *Science* **268**, 1343–1347 (1995).
11. Zinkernagel, R. M., Kreeb, G. & Althage, A. Lymphohemopoietic origin of the immunogenic, virus-antigen-presenting cells triggering anti-viral T-cell responses. *Clin. Immunol. Immunopathol.* **15**, 565–576 (1980).
12. Bevan, M. J. Minor H antigens introduced on H-2 different stimulating cells crossreact at the cytotoxic T cell level during *in vivo* priming. *J. Immunol.* **117**, 2233–2238 (1976).
13. Ren, R. B., Costantini, F., Gorgacz, E. J., Lee, J. J. & Racaniello, V. R. Transgenic mice expressing a human poliovirus receptor: a new model for poliomyelitis. *Cell* **63**, 353–362 (1990).
14. Mandl, S., Sigal, L. J., Rock, K. L. & Andino, R. Poliovirus vaccine vectors elicit antigen-specific cytotoxic T cells and protect mice against lethal challenge with malignant melanoma cells expressing a model antigen. *Proc. Natl Acad. Sci. USA* **95**, 8216–8221 (1998).
15. Ren, R. & Racaniello, V. R. Human poliovirus receptor gene expression and poliovirus tissue tropism in transgenic mice. *J. Virol.* **66**, 296–304 (1992).
16. Bevan, M. J. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. *J. Exp. Med.* **143**, 1283–1288 (1976).
17. Kovacsics-Bankowski, M., Clark, K., Benacerraf, B. & Rock, K. L. Efficient major histocompatibility complex class I presentation of exogenous antigen upon phagocytosis by macrophages. *Proc. Natl Acad. Sci. USA* **90**, 4942–4946 (1993).
18. Kurts, C., Kosaka, H., Carbone, F. R., Miller, J. F. & Heath, W. R. Class I-restricted cross-presentation of exogenous self-antigens leads to deletion of autoreactive CD8(+) T cells. *J. Exp. Med.* **186**, 239–245 (1997).
19. Kurts, C. *et al.* Constitutive class I-restricted exogenous presentation of self antigens *in vivo*. *J. Exp. Med.* **184**, 923–930 (1996).
20. Reis e Sousa, C. & Germain, R. N. Major histocompatibility complex class I presentation of peptides derived from soluble exogenous antigen by a subset of cells engaged in phagocytosis. *J. Exp. Med.* **182**, 841–851 (1995).
21. Rock, K. L. & Clark, K. Analysis of the role of MHC class II presentation in the stimulation of cytotoxic T lymphocytes by antigens targeted into the exogenous antigen–MHC class I presentation pathway. *J. Immunol.* **156**, 3721–3726 (1996).
22. Albert, M. L., Sauter, B. & Bhardwaj, N. Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. *Nature* **392**, 86–89 (1998).
23. Shen, Z., Reznikoff, G., Dranoff, G. & Rock, K. L. Cloned dendritic cells can present exogenous antigens on both MHC class I and class II molecules. *J. Immunol.* **158**, 2723–2730 (1997).
24. Pfeifer, J. D. *et al.* Phagocytic processing of bacterial antigens for class I MHC presentation to T cells. *Nature* **361**, 359–362 (1993).
25. Bachmann, M. F. *et al.* TAP1-independent loading of class I molecules by exogenous viral proteins. *Eur. J. Immunol.* **25**, 1739–1743 (1995).
26. Kovacsics-Bankowski, M. & Rock, K. L. A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* **267**, 243–253 (1995).
27. Ploegh, H. L. Viral strategies of immune evasion. *Science* **280**, 248–253 (1998).
28. Sigal, L. J., Reiser, H. & Rock, K. L. The role of B7-1 and B7-2 costimulation for the generation of CTL responses *in vivo*. *J. Immunol.* **161**, 2740–2745 (1998).
29. Moore, M. W., Carbone, F. R. & Bevan, M. J. Introduction of soluble protein into the class I pathway of antigen processing and presentation. *Cell* **54**, 777–785 (1988).

**Acknowledgements.** We thank R. O. Donis and the members of the Rock laboratory for helpful discussions; R. Welsh, L. Berg, G. Soldevila and N. Kisaiti for critical reading of the manuscript; W. Yong Zang and L. Rothstein for technical assistance; and M. Bevan for EG7 cells, P. Doherty for MC57G cells, and J. Yewdell for vaccinia-OVA and vaccinia SS-SIINFEKL. This work was supported by NIH grants to K.L.R. and R.A. L.J.S. was supported by an NIH Research Training Grant.

Correspondence and requests for materials should be addressed to K.L.R. (e-mail: Kenneth.Rock@banyan.ummed.edu).

## DREAM is a Ca<sup>2+</sup>-regulated transcriptional repressor

Angel M. Carrión, Wolfgang A. Link, Fran Ledo, Britt Mellström & Jose R. Naranjo

Instituto de Neurobiología S. Ramón y Cajal, CSIC, Av. Dr. Arce 37, 28002 Madrid, Spain

Fluxes in amounts of intracellular calcium ions are important determinants of gene expression<sup>1–3</sup>. So far, Ca<sup>2+</sup>-regulated kinases and phosphatases have been implicated in changing the phosphorylation status of key transcription factors and thereby modulating their function<sup>4,5</sup>. In addition, direct effectors of Ca<sup>2+</sup>-induced gene expression have been suggested to exist in the

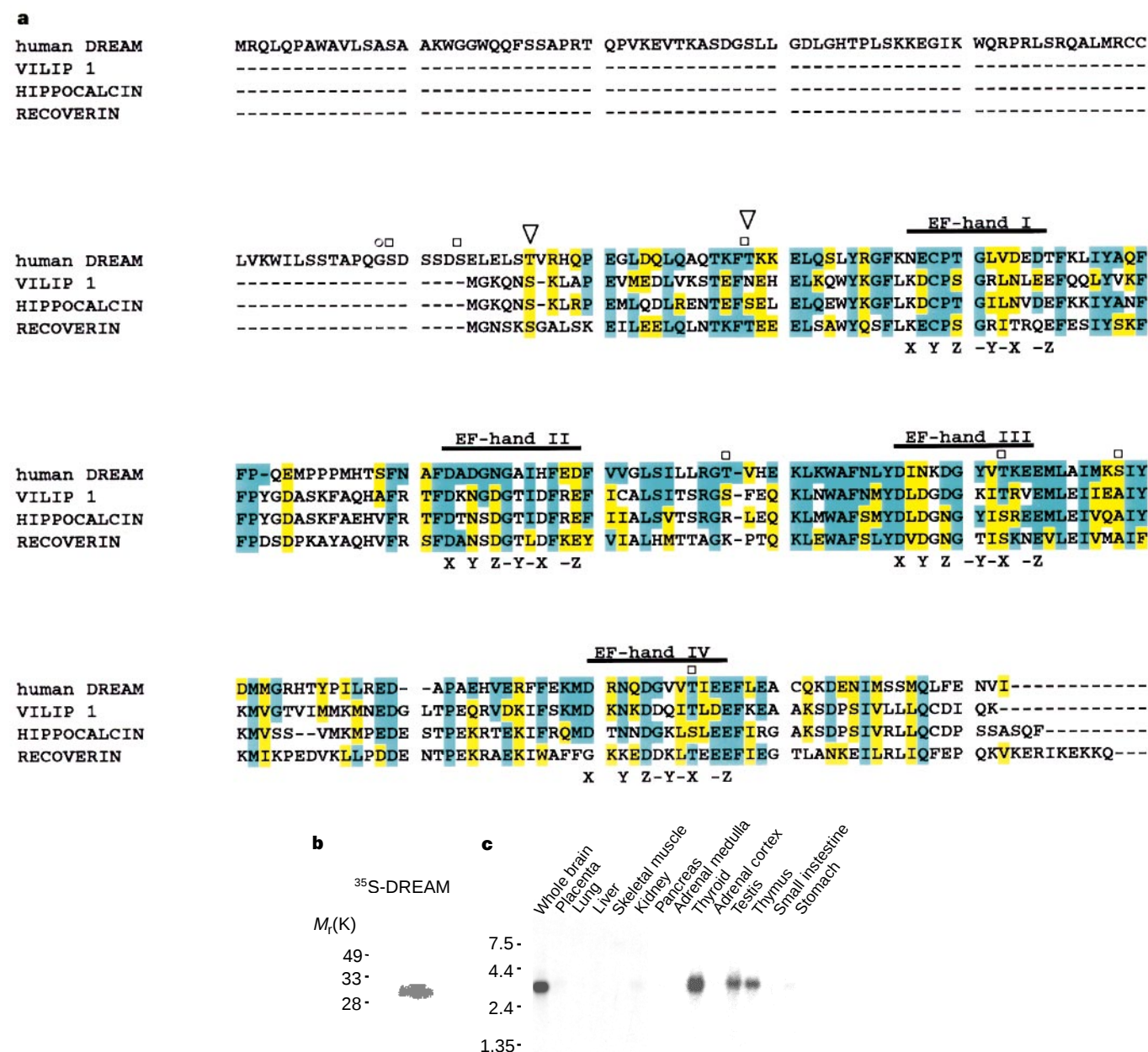


nucleus<sup>2</sup>, although no such effectors have been identified yet. Expression of the human prodynorphin gene, which is involved in memory acquisition and pain<sup>6,7</sup>, is regulated through its downstream regulatory element (DRE) sequence, which acts as a location-dependent gene silencer<sup>8</sup>. Here we isolate a new transcriptional repressor, DRE-antagonist modulator (DREAM), which specifically binds to the DRE. DREAM contains four Ca<sup>2+</sup>-binding domains of the EF-hand type. Upon stimulation by Ca<sup>2+</sup>, DREAM's ability to bind to the DRE and its repressor function are prevented. Mutation of the EF-hands abolishes the response of DREAM to Ca<sup>2+</sup>. In addition to the prodynorphin promoter, DREAM represses transcription from the early response gene *c-fos*. Thus, DREAM represents the first known Ca<sup>2+</sup>-binding protein to function as a DNA-binding transcriptional regulator.

Expression of the human prodynorphin gene is regulated by

derepression at the DRE site<sup>8</sup>. To isolate complementary DNAs encoding proteins able to bind to the DRE sequence, we screened a human caudate expression library using a double-stranded DRE oligonucleotide as a probe. One positive cDNA clone encoding a protein of 284 amino acids (Fig. 1a), with a predicted relative molecular mass of 31,800 (*M<sub>r</sub>* 31.8K), was obtained. A protein of this size was consistently translated from the cDNA using reticulocyte lysates (Fig. 1b). As the DRE site functions as a silencer<sup>8</sup> we named the DRE-binding protein DRE-antagonist modulator (DREAM). Northern blot analysis in human tissues revealed a 3.0-kilobase DREAM transcript that was highly expressed in brain, thymus, thyroid gland and testis (Fig. 1c).

We next confirmed that the 31.8K DREAM protein specifically binds to the DRE. Crude extracts were prepared from *Escherichia coli* transformed with a DREAM cDNA expression construct and analysed by electrophoretic mobility-shift analysis (EMSA) using

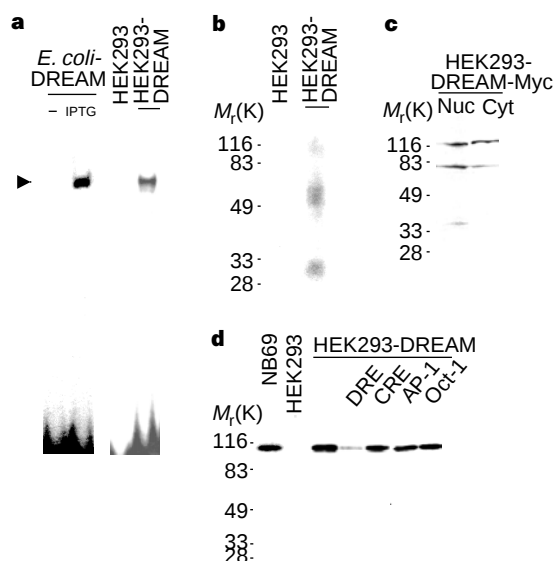


**Figure 1** The DREAM gene products. **a**, Amino-acid sequence of human DREAM and alignment with other Ca<sup>2+</sup>-binding proteins. Conserved amino acids are coloured blue and conservative changes are coloured yellow. The positions of the X, Y, Z, -Y, -X and -Z Ca<sup>2+</sup>-chelating residues containing oxygen in their side chains<sup>9</sup> are indicated within each EF-hand. Potential residues for myristoylation

(circles) and for phosphorylation by protein kinase C (triangles) or casein kinase II (squares) are also indicated. **b**, *In vitro*-translated <sup>35</sup>S-labelled DREAM protein from reticulocyte lysates. **c**, Northern blot analysis of DREAM messenger RNA in human tissues. RNA size markers (in kilobases) are shown to the left.

the DRE as probe. A DRE-retarded band was observed in bacterial extracts only after induction with isopropyl- $\beta$ -D-thiogalactoside (IPTG) (Fig. 2a). The DRE-retarded band was also observed using nuclear extracts from HEK293 cells stably transfected with DREAM (HEK293-DREAM), but no band was present in nuclear extracts from wild-type HEK293 cells (Fig. 2a). Ultraviolet-light-induced crosslinking of the DRE probe to bacterial extracts or nuclear extracts from HEK293-DREAM cells reproducibly resulted in the appearance of three bands, with apparent molecular masses of 31K, 60K and 110K, after SDS-polyacrylamide gel electrophoresis (Fig. 2b and data not shown). Again, no bands were observed with wild-type HEK293 cells or with non-induced bacterial extracts (Fig. 2b and data not shown). These results indicate that DREAM may bind to the DRE as a monomer, dimer and tetramer.

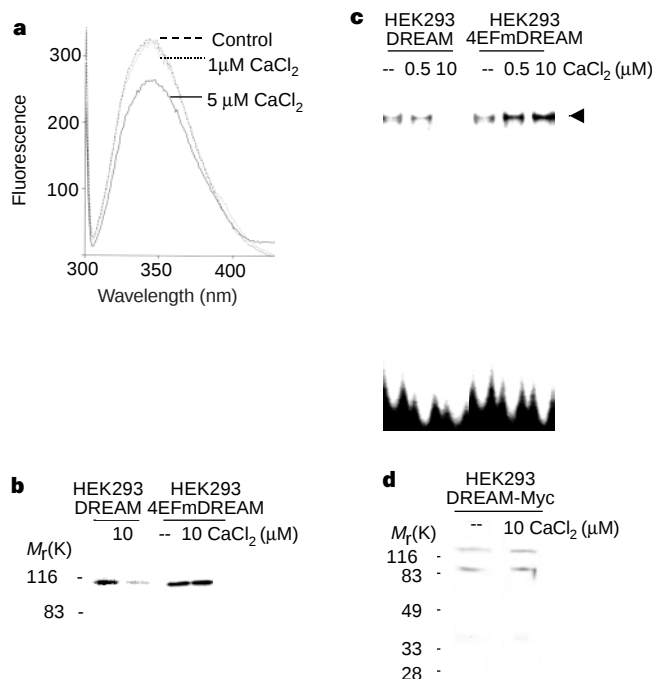
Western blot analysis of purified nuclei and cytosolic fractions prepared from stably transfected HEK293 cells expressing a DREAM fusion protein containing a Myc tag (predicted  $M_r$  35.6K) revealed three bands corresponding to monomeric, dimeric and tetrameric DREAM-Myc in the nuclear extracts and dimeric and tetrameric DREAM-Myc in the cytosolic extracts (Fig. 2c). Whether the presence of DREAM in the cytosol is simply the result of overexpression or whether it has functional significance is unknown at present. When nuclear extracts from HEK293-DREAM cells were analysed by southwestern blotting, only a 110K specific band was reproducibly detected (Fig. 2d). This agrees with previous results obtained using NB69 cells, a human neuroblastoma cell line containing endogenous DRE-binding activity<sup>8</sup>. Taken together, these results indicate that the 31.8K DREAM protein binds with high affinity to the DRE sequence as a homomultimer but is still able to bind to the DRE as a monomer, although with lower affinity.



**Figure 2** DREAM binds to the DRE. **a**, Electrophoretic mobility shift assay showing the specific DRE-retarded band (arrowhead) after overexpression of DREAM in *E. coli* (*E. coli*-DREAM) and HEK293 cells (HEK293-DREAM), IPTG, isopropyl- $\beta$ -D-thiogalactoside. **b**, Ultraviolet-crosslinking analysis of the DRE probe and nuclear extracts from HEK293 cells overexpressing DREAM. Monomeric, dimeric and tetrameric DREAM proteins are observed. **c**, Western blot analysis of nuclear (Nuc) and cytosolic (Cyt) extracts of HEK293 cells overexpressing a DREAM-Myc fusion protein. **d**, Southwestern analysis of nuclear extracts from HEK293-DREAM cells, showing the 110K DREAM-DRE complex. The 110K band obtained with the endogenous DREAM in nuclear extracts from NB69 cells is shown for comparison. Competition by cold oligonucleotides containing DRE, cAMP-responsive element (CRE), AP-1 or Oct-1 sites is also shown.

Analysis of the DREAM protein revealed the presence of four  $\text{Ca}^{2+}$ -binding domains known as EF-hands<sup>9</sup>. Binding to  $\text{Ca}^{2+}$  within each EF-hand is directed by six highly conserved residues (X, Y, Z, -Y, -X and -Z; Fig. 1a) located at positions 1, 3, 5, 7, 9 and 12 within a 12-amino-acid loop that is flanked by two  $\alpha$ -helices (the 12-amino-acid loop and the two  $\alpha$ -helices together constitute the EF-hand)<sup>9</sup>. The second, third and fourth EF-hands of DREAM strictly conform to the EF-hand consensus<sup>9</sup>, whereas the first EF-hand is less well conserved. Database searches revealed similarity between DREAM and the recoverin-like  $\text{Ca}^{2+}$ -binding proteins, vilip1, recoverin and hippocalcin<sup>9,10</sup>. Alignment of human DREAM with these  $\text{Ca}^{2+}$ -binding proteins shows that the similarity is especially high within the regions containing the EF-hands (Fig. 1a). In addition, analysis of the DREAM sequence predicts several consensus sites for phosphorylation by protein kinase C and casein kinase II and a site for myristoylation (Fig. 1a).

To determine the functional significance of the EF-hands present in DREAM we analysed the tryptophan fluorescence emission spectra of DREAM in the presence and absence of  $\text{Ca}^{2+}$ . Addition of  $5 \mu\text{M}$   $\text{Ca}^{2+}$  decreased the fluorescence intensity (Fig. 3a), indicating that DREAM specifically binds  $\text{Ca}^{2+}$  at micromolar concentrations in solution. Using site-directed mutagenesis we prepared DREAM proteins with mutations in one, two or three of the EF-hands, namely 2, 3, 4, 2/3, 2/4, 3/4 and 2/3/4EFmutDREAM, in which residues X and Y (Fig. 1a) responsible for binding to  $\text{Ca}^{2+}$  within the second, third and/or fourth EF-hands<sup>9</sup> were substituted with alanine. Results obtained with 4EFmutDREAM are shown (Fig. 3b, c). Analysis of nuclear extracts from HEK293 cells overexpressing DREAM or 4EFmutDREAM revealed the presence of equivalent 110K bands (Fig. 3b; southwestern blotting) and DRE-retarded



**Figure 3** DREAM is a  $\text{Ca}^{2+}$ -binding protein. **a**, Fluorescence emission spectra of  $0.5 \mu\text{M}$  DREAM in the absence of  $\text{Ca}^{2+}$  (dashed line) and in the presence of  $1 \mu\text{M}$  (dotted line) or  $5 \mu\text{M}$  (solid line)  $\text{Ca}^{2+}$ . (The fluorescence is in arbitrary units.) **b**, Southwestern analysis of HEK293 cells after stable overexpression of DREAM or 4EFmutDREAM. The differential effect of  $\text{Ca}^{2+}$  on the appearance of the 110K DREAM and 4EFmutDREAM bands is shown. **c**, Electrophoretic mobility shift assay showing the lack of effect of  $\text{Ca}^{2+}$  on the DRE-retarded band obtained with the 4EFmutDREAM protein. For comparison, the blockade by  $\text{Ca}^{2+}$  of the DRE-retarded band obtained with the wild-type DREAM protein is shown. **d**, Western blot analysis of nuclear DREAM-Myc in the absence and presence of calcium.

bands (Fig. 3c; EMSA). However, although addition of  $10 \mu\text{M}$   $\text{Ca}^{2+}$  to the nuclear extracts from HEK293-DREAM cells significantly reduced the intensity of the 110K band (Fig. 3b) and the DRE-retarded band (Fig. 3c),  $\text{Ca}^{2+}$  completely failed to affect either the 110K band (Fig. 3b) or the DRE-retarded band (Fig. 3c) in nuclear extracts from HEK293-4EFmutDREAM cells. Comparable results were obtained using the other EF-hand DREAM mutants (data not shown). Western blot analysis of nuclear extracts from HEK293-DREAM-Myc cells in the absence or presence of  $\text{Ca}^{2+}$  showed an

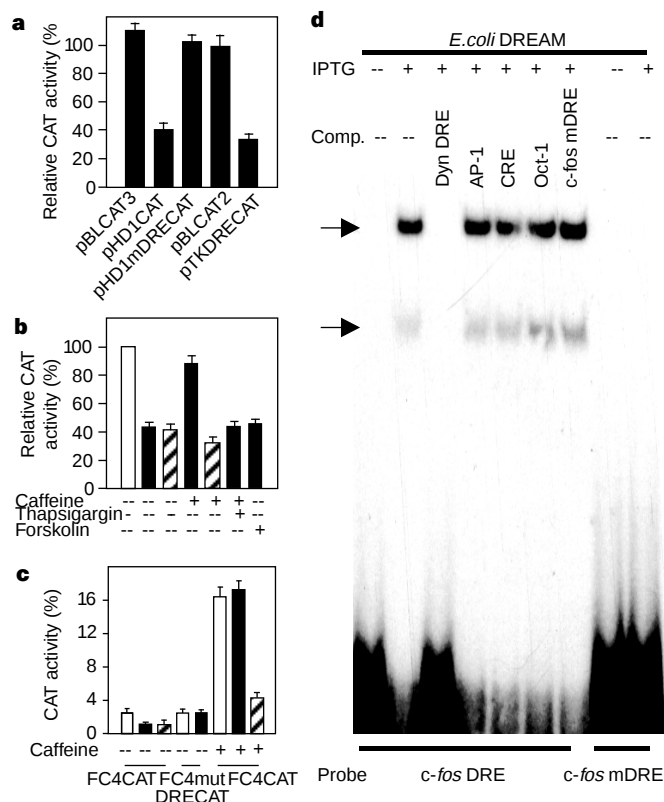
identical pattern of monomer, dimer and tetramer bands (Fig. 3d). Thus,  $\text{Ca}^{2+}$  binding to the EF-hand induces a conformational change in DREAM that precludes its binding to the DRE without affecting the stability of tetrameric DREAM.

To study the DRE-dependent transcriptional-regulatory function of DREAM, we performed transient transfection assays with the prodynorphin promoter (reporter plasmid pHD1CAT) or the heterologous promoter construct pTKDRECAT<sup>8</sup>, together with a DREAM expression vector. Co-transfection with DREAM significantly repressed basal transcription from both pHD1CAT and pTKDRECAT in HEK293 cells (Fig. 4a) but did not modify basal transcription from the empty reporters pBLCAT3 and pBLCAT2 (Fig. 4a). Furthermore, DREAM failed to repress transcription from reporter plasmid pHD1mDRECAT<sup>8</sup>, in which the DRE site is mutated and fails to bind to DREAM. These results indicate that DREAM functions as a transcriptional repressor through binding to the DRE.

We next explored the effect of  $\text{Ca}^{2+}$  on DREAM-mediated transcriptional repression. We compared the transcriptional responses of DREAM and the DREAM EF-hand mutants to intracellular  $\text{Ca}^{2+}$  release induced by caffeine treatment<sup>11</sup>. Co-transfection with DREAM or 4EFmutDREAM resulted in similar repression of the reporter function under basal conditions (Fig. 4b), consistent with the finding that 4EFmutDREAM is still able to bind to the DRE (Fig. 3b, c). However, whereas increased intracellular  $\text{Ca}^{2+}$  completely prevented the repressive effect of DREAM (Fig. 4b), caffeine did not modify the repression exerted by the 4EFmutDREAM because of its inability to sense the  $\text{Ca}^{2+}$  rise. The same results were obtained with the other single, double and triple DREAM EF-hand mutants (data not shown), indicating that any of the single mutations alone is enough to preclude the effect of stimulation by  $\text{Ca}^{2+}$ . The derepression of DREAM by caffeine is related to the release of  $\text{Ca}^{2+}$ : it was blocked by thapsigargin and ryanodine, drugs that block intracellular  $\text{Ca}^{2+}$  release<sup>12</sup> (Fig. 4b and data not shown). Moreover, the depression by caffeine is not related to phosphodiesterase inhibition and subsequent increases in cyclic AMP levels, as direct activation of protein kinase A by forskolin failed to affect the repression of pTKDRECAT by DREAM in HEK293 cells (Fig. 4b).

We next wondered whether DREAM could repress transcription of other  $\text{Ca}^{2+}$ -activated genes. We studied the *c-fos* gene, which is known to be  $\text{Ca}^{2+}$ -regulated<sup>1,3,13</sup>. Upon transient co-transfection in HEK293 cells, basal expression from FC4CAT, a reporter containing the *c-fos* promoter<sup>14</sup>, was reduced by DREAM (Fig. 4c). Sequence analysis of the *c-fos* promoter revealed the presence of the putative DRE sequence, GNARYYRAG, downstream of the TATA box. To determine whether this sequence is responsible for DREAM-mediated repression, we mutated the *c-fos* DRE within reporter plasmid FC4CAT, generating reporter FC4mDRECAT. Transient co-transfection in HEK293 cells showed that transcription from FC4mDRECAT was not repressed by DREAM (Fig. 4c). Using extracts from bacteria or HEK293 cells overexpressing DREAM for EMSA with the *c-fos* DRE, we obtained two DRE-retarded bands (Fig. 4d and data not shown). An oligonucleotide containing the dynorphin DRE, but not the mutated *c-fos* DRE or unrelated AP-1, cAMP responsive element or Oct-1 sites, competed with the *c-fos* DRE in EMSAs (Fig. 4d). These results indicate that DREAM tonically represses *c-fos* by acting at the *c-fos* DRE site. Interestingly, activation of transcription of the *c-fos* reporter by intracellular  $\text{Ca}^{2+}$  release was blocked by the 4EFmutDREAM mutant, and DREAM did not repress transcription of this reporter in the presence of  $\text{Ca}^{2+}$  (Fig. 4c).

Here we have shown for the first time, to our knowledge, that a nuclear  $\text{Ca}^{2+}$  sensor, DREAM, directly represses transcription by binding specifically to DNA. DREAM functions as a homotetramer and its affinity for DNA is reduced upon binding to  $\text{Ca}^{2+}$ . This represents a direct mechanism for  $\text{Ca}^{2+}$ -induced gene expression that is not dependent on changes in the activity of other transcriptional



**Figure 4** DREAM is a  $\text{Ca}^{2+}$ -dependent transcriptional repressor. **a**, Repression by DREAM of the DRE-containing reporters pHD1CAT and pTKDRECAT after transient transfection of these reporters with DREAM into HEK293 cells. The lack of repression of pHD1mutDRECAT is also shown. pBLCAT3 and pBLCAT2 are empty reporter plasmids. Each bar represents the ratio of CAT activity from each construct co-transfected with DREAM or the empty expression vector pCDNA3. **b**, Repression by DREAM (black bars) or 4EFmutDREAM (hatched bars) of basal transcription from pTKDRECAT after transient co-transfection into HEK293 cells. DREAM's repressor activity is blocked upon stimulation with  $\text{Ca}^{2+}$  (caffeine; 10 mM), whereas caffeine does not block repression mediated by 4EFmutDREAM. Forskolin treatment (25  $\mu\text{M}$ ) failed to block repression by DREAM of the TKDRECAT reporter whereas thapsigargin (1  $\mu\text{M}$ ) completely blocked the derepression by caffeine. Bars represent the amount of CAT activity produced from the TKDRECAT reporter transfected together with DREAM or 4EFmutDREAM, compared with the activity from the TKDRECAT reporter transfected together with pCDNA3 (open bar) for each experimental condition. **c**, DREAM represses basal transcription of the *c-fos* gene. Transient transfection experiments in HEK293 cells show the repression by DREAM (black bars) of basal transcription from the FC4CAT reporter but not from the FC4mDRECAT reporter. Transactivation of the FC4CAT reporter by caffeine in HEK293 cells is blocked by expression of the 4EFmutDREAM mutant (hatched bar) but not by DREAM (black bar). Empty bars represent co-transfection with empty expression vector pCDNA3. **d**, Electrophoretic mobility shift assay, using the *c-fos* DRE as a probe and bacterially expressed DREAM. The specific *c-fos*-DRE-retarded bands (arrows) are specifically competed (Comp) by cold dynorphin (Dyn) DRE but not by mutated *c-fos* mDRE, AP-1, CRE or Oct-1. Moreover, the mutated *c-fos* mDRE failed to form the *c-fos*-DRE-retarded bands when used as a probe.



effectors through phosphorylation<sup>4,15,16</sup> or protein–protein interactions<sup>17</sup>. Target genes for DREAM repression include prodynorphin and *c-fos*, indicating that DREAM may have a broad role in the regulation of a range of genes. Future studies addressing the functional consequences of tissue-specific suppression of DREAM function should therefore be an important contribution to our understanding of how calcium differentially regulates gene expression. □

## Methods

**Library screening and RNA analysis.** A double-stranded oligonucleotide, 5'-CAAGCCGGAGTCAAGGAGGCCCTGCCGGAGTCAAGGAGGCCCTG-3', containing two tandemly repeated DREs was used to screen a human caudate expression library (Clontech). Northern blots of human tissues were purchased from Clontech.

**Cell culture, transfection and chloramphenicol acetyltransferase (CAT) analysis.** The cells were grown in DMEM medium supplemented with 10% fetal calf serum, 2 mM glutamine and 50 µg ml<sup>-1</sup> gentamycin. Transfections by calcium phosphate precipitation and CAT activity assays were done as described<sup>8</sup>. Treatments with caffeine or forskolin were performed 6 h before collecting the cells. When used, thapsigargin (1 µM) or ryanodine (100 µM) was added to the cells 30 min before caffeine.

**Plasmid constructions and site-directed mutagenesis.** The reporter plasmids pHD1CAT, pHD1mDRECAT and pTKDRECAT have been described previously<sup>8</sup>. Mutations were prepared using *Pfu* DNA polymerase (Quik-Change, Stratagen). The oligonucleotide 5'-CTTCGAGAAATGGCCCGGAT-CCAGGATGGGGTAG-3' was used for the mutation of the DREAM fourth EF-hand. To mutate the FC4CAT reporter we used the oligonucleotide *c-fos* mutDRE, 5'-CTGCAGCGAGCAACAAAGAATCCAAGAC-3'. The DREAM cDNA was cloned in the prokaryotic expression vector pTrcHis and the overexpressed protein was affinity-purified according to the suggested protocol (Invitrogen). For overexpression of DREAM and DREAM-Myc in HEK293 cells, we cloned the cDNA in pCDNA3 and pCDNA3.1, respectively.

**Protein analysis.** EMSA, ultraviolet crosslinking and southwestern analysis were performed as described<sup>8</sup>. The *c-fos* DRE oligonucleotide was 5'-CTGCAGCGAGCAACTGAGAATCCAAGAC-3'. DNA-competition experiments were performed as described<sup>8</sup>. Addition of Ca<sup>2+</sup> to the nuclear extracts, where indicated, was done at room temperature 20 min before loading in the gel. Tryptophan fluorescence spectra of bacterially expressed DREAM was performed using a Shimadzu RF-5001 PC fluorimeter nearly as described<sup>10</sup>. Subcellular fractionation was done as described<sup>18</sup>. Nuclear and cytosolic fractions were electrophoresed in a 12% SDS polyacrylamide gel, transferred to poly(vinylidene) fluoride membrane and incubated with anti-Myc monoclonal antibody (1:5,000, Invitrogen). The blots were developed with biotin-conjugated anti-mouse antibody (1:10,000, Jackson) and peroxidase-conjugated streptavidin (1:1,000, Amersham). Detection was by enhanced chemiluminescence (Amersham).

Received 11 November 1998; accepted 11 January 1999.

- Bading, H., Ginty, D. D. & Greenberg, M. E. Regulation of gene expression in hippocampal neurons by distinct calcium signaling pathways. *Science* **260**, 181–186 (1993).
- Ghosh, A. & Greenberg, M. E. Calcium signaling in neurons: molecular mechanisms and cellular consequences. *Science* **268**, 239–247 (1995).
- Hardingham, G. E., Chawla, S., Johnson, C. M. & Bading, H. Distinct functions of nuclear and cytoplasmic calcium in the control of gene expression. *Nature* **385**, 260–265 (1997).
- Sheng, M., Thompson, M. A. & Greenberg, M. E. CREB: a Ca<sup>2+</sup>-regulated transcription factor phosphorylated by calmodulin-dependent kinases. *Science* **252**, 1427–1430 (1991).
- Dolmetsch, R. E., Lewis, R. S., Goodnow, C. C. & Healy, J. I. Differential activation of transcription factors induced by Ca<sup>2+</sup>-response amplitude and duration. *Nature* **386**, 855–858 (1997).
- Weiskopf, M. G., Zalutsky, R. A. & Nicoll, R. A. The opioid peptide dynorphin mediates heterosynaptic depression of hippocampal mossy fiber synapses and modulates long-term potentiation. *Nature* **362**, 423–427 (1993).
- Naranjo, J. R., Mellström, B., Achaval, M. & Sassone-Corsi, P. Molecular pathways of pain: Fos/Jun-mediated activation of the prodynorphin gene through a noncanonical AP-1 site. *Neuron* **6**, 607–617 (1991).
- Carrión, A. M., Mellström, B. & Naranjo, J. R. PKA-dependent derepression of the human prodynorphin gene via differential binding to an intragenic silence element. *Mol. Cell. Biol.* **18**, 6921–6929 (1998).
- Nakayama, S. & Kretsinger, R. H. Evolution of the EF-hand family of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 473–507 (1994).
- Dizhoor, A. M. *et al.* Recoverin: a calcium sensitive activator of retinal rod guanylate cyclase. *Science* **251**, 915–918 (1991).
- Hernández-Cruz, A., Salsa, F. & Adams, P. R. Subcellular calcium transients visualized by confocal microscopy in a voltage-clamped vertebrate neuron. *Science* **147**, 858–862 (1990).

- Ehrlich, B. E., Kaftan, E., Bezprozvannaya, S. & Bezprozvannaya, I. The pharmacology of intracellular Ca<sup>2+</sup>-release channels. *Trends Pharmacol. Sci.* **15**, 145–149 (1994).
- Morgan, J. I. & Curran, T. Role of ion flux in the control of *c-fos* expression. *Nature* **322**, 552–555 (1986).
- Sassone-Corsi, P., Sisson, J. C. & Verma, I. M. Transcriptional autoregulation of the proto-oncogene *fos*. *Nature* **334**, 314–319 (1988).
- Deisseroth, K., Heist, E. K. & Tsien, R. W. Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons. *Nature* **392**, 198–202 (1998).
- Johnson, C. H., Hill, C. S., Clawla, S., Treisman, R. & Bading, H. Calcium controls gene expression via three distinct pathways that can function independently of the Ras/mitogen activated protein kinase (ERKs) signaling cascade. *J. Neurosci.* **17**, 6189–6202 (1997).
- Corneliusson, B. *et al.* Calcium/calmodulin inhibition of basic-helix-loop-helix transcription factor domains. *Nature* **368**, 760–764 (1994).
- Stegh, A. *et al.* DEDD, a novel death effector domain-containing protein, targeted to the nucleolus. *EMBO J.* **17**, 5974–5986 (1998).

**Acknowledgements.** We thank N. S. Foulkes for critical reading of the manuscript and I. DomPablo and D. Campos for technical assistance. Work in this laboratory is supported by grants from DGICYT, CAM, Europharma SA and Janssen-Cilag SA (to J.R.N.).

Correspondence and requests for material should be addressed to J.R.N. (e-mail: jrnaranjo@samba.cnb.uam.es). The sequence of human DREAM cDNA has been deposited at the EMBL DNA Databank under accession number AJ131730.

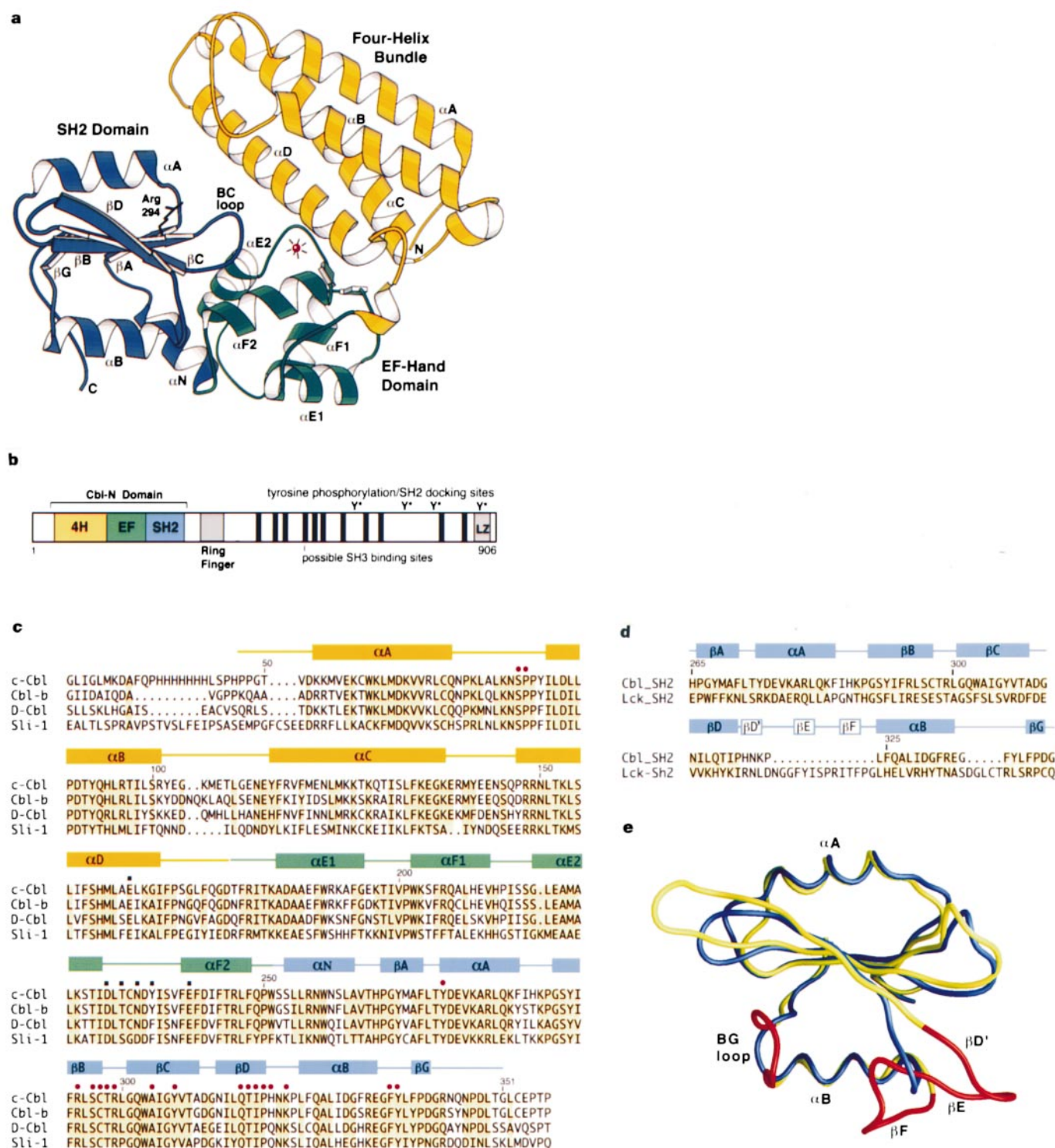
## Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase

Wuyi Meng\*, Sansana Sawasdikosol†, Steven J. Burakoff† & Michael J. Eck\*

Departments of \*Biological Chemistry and Molecular Pharmacology and †Pediatrics, Harvard Medical School, and Departments of \*Cancer Biology and †Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

Cbl is an adaptor protein that functions as a negative regulator of many signalling pathways that start from receptors at the cell surface<sup>1–4</sup>. The evolutionarily conserved amino-terminal region of Cbl (Cbl-N) binds to phosphorylated tyrosine residues and has cell-transforming activity. Point mutations in Cbl that disrupt its recognition of phosphotyrosine also interfere with its negative regulatory function and, in the case of v-cbl, with its oncogenic potential<sup>5</sup>. In T cells, Cbl-N binds to the tyrosine-phosphorylated inhibitory site of the protein tyrosine kinase ZAP-70<sup>6</sup>. Here we describe the crystal structure of Cbl-N, both alone and in complex with a phosphopeptide that represents its binding site in ZAP-70. The structures show that Cbl-N is composed of three interacting domains: a four-helix bundle (4H), an EF-hand<sup>7</sup> calcium-binding domain, and a divergent SH2 domain<sup>8</sup> that was not recognizable from the amino-acid sequence of the protein. The calcium-bound EF hand wedges between the 4H and SH2 domains and roughly determines their relative orientation. In the ligand-occupied structure, the 4H domain packs against the SH2 domain and completes its phosphotyrosine-recognition pocket. Disruption of this binding to ZAP-70 as a result of structure-based mutations in the 4H, EF-hand and SH2 domains confirms that the three domains together form an integrated phosphoprotein-recognition module.

Cbl becomes tyrosine-phosphorylated upon engagement of several cell-surface receptors, including the multichain immune receptors, and growth-factor and cytokine receptors<sup>1–4</sup>. The amino-acid sequence of Cbl reveals a RING-finger<sup>9</sup> domain adjacent to the phosphotyrosine-binding Cbl-N segment, and a carboxy-terminal region with numerous docking sites for SH3- and SH2-containing proteins (Fig. 1). The oncogenic v-Cbl includes only the first 357 residues of Cbl and is a potent transforming protein<sup>10</sup>. The highly conserved Cbl-N region binds to the receptor for epidermal growth factor (EGFR)<sup>11</sup>, Syk<sup>12</sup> and the negative-regulatory phosphorylation



**Figure 1** Cbl domain structure and sequence comparisons. **a**, Ribbon diagram of unliganded Cbl-N. The N-terminal 4H domain is coloured yellow, the EF-hand domain green, and the SH2 domain blue. Secondary-structure elements are labelled  $\alpha A$ – $\alpha D$  in the 4H domain and by established conventions for the EF-hand and SH2 domains. The bound  $Ca^{2+}$  ion is indicated by a red sphere. Arginine 294 is universally conserved in SH2 domains and participates in phosphotyrosine coordination. **b**, Diagram of c-Cbl domain structure. The Cbl-N region and adjacent RING finger domain are conserved in all Cbl homologues. The C-terminal region, which contains proline-rich segments and tyrosine phosphorylation sites, is more variable and is completely absent in D-Cbl. A putative leucine zipper has been found near the C terminus of Cbl. **c**, Aligned sequences of the Cbl-N portion of human c-Cbl, human Cbl-b, *Drosophila* D-Cbl, and Sli-1.

Residues that are identical in at least three of the sequences are shaded yellow. Secondary-structure elements are shown above the sequence and are coloured as in **a** and **b**. Black squares indicate residues that coordinate calcium. Red circles mark residues that interact with the bound ZAP-70 peptide. **d**, Structure-based sequence alignment of Cbl and Lck<sup>23</sup> SH2 domains. Seventy structurally equivalent residues are shaded yellow;  $\alpha$ -carbons of these seventy residues superimpose with an r.m.s.d. of 1.47 Å. The secondary-structure elements that are present in Lck and other SH2 domains, but not in the Cbl SH2 domain, are indicated by open boxes. **e**, Superposition of the Cbl SH2 domain (blue) with the Lck SH2 domain (yellow). The structural elements that are absent in the Cbl domain are red.



site in ZAP-70<sup>6</sup>. Genetic studies in *Caenorhabditis elegans* revealed that the Cbl homologue Sli-1 inhibits vulval induction by the EGFR<sup>13</sup>. Expression of *D-Cbl*, a *Drosophila melanogaster* Cbl homologue, disrupts EGFR-regulated development of the R7 photoreceptor<sup>14</sup>. Cbl also diminishes FcεRI-mediated degranulation in mast cells by inhibiting the tyrosine kinase Syk<sup>15</sup>. Mutational analysis demonstrates that Cbl-N is central to these functions. To understand better the diverse recognition and regulatory functions of Cbl-N, we have determined its three-dimensional structure.

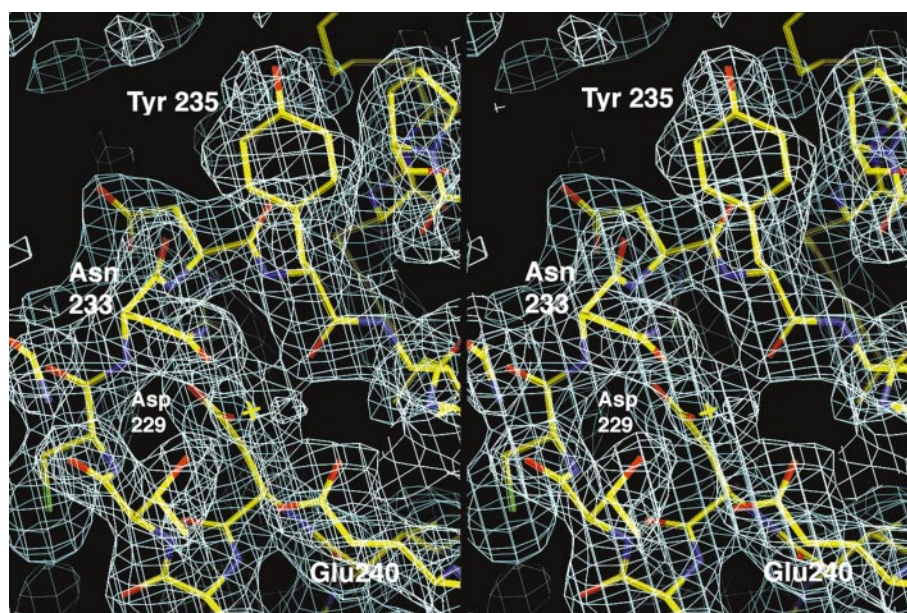
As shown in Fig. 1, Cbl-N comprises three interacting domains: an N-terminal four-helix bundle (4H), a calcium-binding domain with the EF-hand fold, and an unusual SH2 domain. None of these folding motifs were previously recognized in the amino-acid sequence of Cbl. In spite of the structural and functional similarity of this unusual SH2 domain with other SH2 domains, it shares very little sequence identity (~11%) with them.

The N-terminal 4H domain contains four long α-helices. Structural comparisons with the DALI<sup>16</sup> server show that it has a topology and overall structure similar to many functionally unrelated four-helical proteins, including cytochrome *c*, interleukin-5 and apolipoprotein III. The C and D helices in this domain pack against the adjacent EF-hand domain, and a highly conserved loop connecting the A and B helices contacts the SH2 domain.

The EF-hand motif is similar to those of classical EF-hand proteins such as troponin C and calmodulin. The classical EF-hand fold contains two calcium-binding sites that are formed by the loops connecting the two pairs of E and F helices<sup>7</sup>. In Cbl, we observe a bound Ca<sup>2+</sup> ion only in the more carboxy-terminal E2F2 motif. Although it lacks a canonical calcium-binding sequence pattern, the loop coordinates calcium in a pentagonal-bipyramidal fashion, as in canonical EF-hand proteins (Fig. 2). The paucity of acidic residues in the Cbl loop may be compensated for by interactions with the four-helix bundle, which contributes one of the axial calcium ligands. Glu 164 in helix D coordinates calcium through a bridging water molecule. Typical EF hands use an acidic residue at position 9 in the EF loop to coordinate at this position, sometimes through a bridging water molecule<sup>17</sup>. The E1F1 loop in Cbl apparently cannot bind calcium: it is one residue shorter than a canonical loop, lacks conserved acidic residues, and does not bind calcium in either of our structures. EF-hand domains occur in

several multidomain signalling proteins, often without a clear regulatory function. Phospholipase Cδ contains a calcium-bound EF hand, but it does not interact strongly with other domains in the protein and its function is unknown<sup>18</sup>. Eps15-homology (EH) domains are essentially EF-hand proteins; some of these domains bind calcium, but not in a way that affects their binding with ligands<sup>19</sup>. Structures determined for STAT proteins<sup>20,21</sup> have revealed an EF-hand-like 'linker' domain that makes substantial contact with an adjacent SH2 domain; however, the EF/SH2 interaction in these STAT structures is unlike that in Cbl and the STAT domains do not bind calcium.

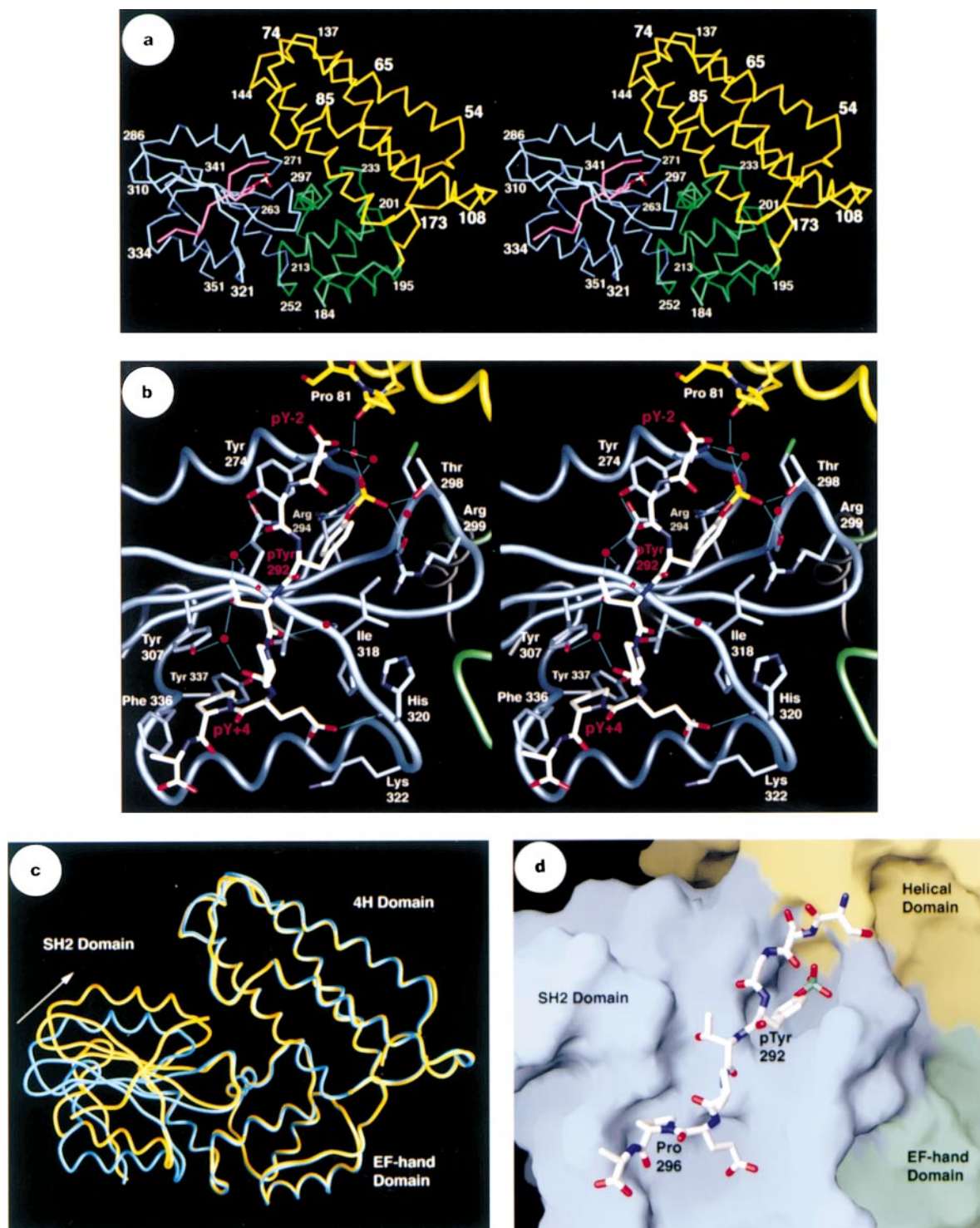
The SH2 domain in Cbl-N retains the general helix-sheet-helix architecture of the SH2 fold (Fig. 1 d,e)<sup>8</sup>, but lacks the secondary β-sheet, comprising β-strands D', E and F, and also a prominent BG loop. The BG and EF loops in most SH2 domains form a pocket or groove that binds the specificity-determining residues just C-terminal to phosphotyrosine (pY). The phosphotyrosine-binding pocket in Cbl-N is better conserved. The universally conserved arginine (Arg βB5), which makes two hydrogen bonds with the phosphate group, is present in Cbl (Arg 294). In spite of the considerable divergence of the sequence and structure of the Cbl SH2 domain from other SH2 domains, the ZAP-70 phosphopeptide binds as other SH2/phosphopeptide complexes do: the peptide extends across the surface of the domain, roughly perpendicular to the edge of the central β-sheet. The phosphotyrosine residue inserts into a pocket on one side of the sheet, and C-terminal residues are coordinated on the opposite side (Fig. 3). A portion of the phosphotyrosine pocket is formed by conserved residues in the AB loop of the 4H domain (Fig. 3d). The carbonyl of Pro 81 hydrogen-bonds to a phosphate oxygen through a bridging water molecule, and Pro 82, which is in a *cis* conformation, packs against the phosphotyrosine-binding BC loop. Additional interactions within the phosphotyrosine pocket are similar to those in other SH2-domain complexes (Fig. 3b). Comparison of the liganded and unliganded structures shows that phosphopeptide binding induces a 'closure' of the domains, which creates the intimate association of the 4H domain with the SH2 domain (Fig. 3c). The relative orientations of the 4H domain and EF-hand domain are unchanged, but the SH2 domain rotates by more than 10°, shifting its position by ~5 Å. The shifted domain packs against helix D and



**Figure 2** Stereo diagram showing the experimental electron density map in the region of the calcium-binding site. The map is a three-fold-averaged SIRAS map, calculated with phases extended to 2.25 Å. The map is contoured at 1.2 σ, and is

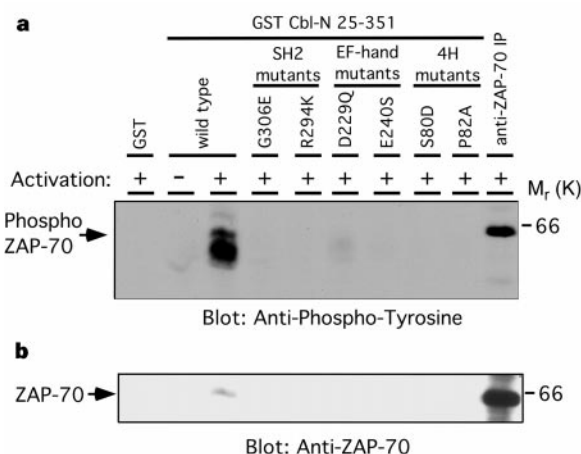
shown with the refined atomic model. A yellow cross marks the position of the bound calcium ion. The observed pentagonal-bipyramidal coordination is characteristic of EF-hand proteins.





**Figure 3** Structure of the Cbl-N / ZAP-70 pY292 complex. **a**, Stereo diagram showing an  $\alpha$ -carbon trace of the complex. The bound ZAP-70 phosphopeptide is shown in magenta. **b**, Stereo diagram showing the interactions with the ZAP-70 phosphopeptide. The bound peptide is shown in white. Red spheres represent ordered water molecules that bridge Cbl-N and the bound peptide. Thin blue lines represent hydrogen bonds. In the phosphotyrosine pocket, Tyr 274 in Cbl makes an 'edge-face' interaction with the phosphotyrosine ring, and its hydroxyl group hydrogen-bonds to the carbonyl oxygen of Gly 291 in the ZAP-70 peptide. An arginine residue found in this position in most SH2 domains makes an 'amino-aromatic' interaction with the phosphotyrosine ring and also hydrogen-bonds with the carbonyl of the pY-1 residue of the bound peptide<sup>8</sup>. C-terminal to the phosphotyrosine, the proline at position pY+4 in the ZAP-70 peptide binds in a hydrophobic cleft formed by Tyr 307, Phe 336 and Tyr 337, and the glutamic acid residue at pY+3 hydrogen-bonds with the backbone amide of His 320. **c**, Superposition of the liganded (yellow) and unliganded (blue) Cbl-N structures reveals a shift in the

position of the SH2 domain upon phosphopeptide binding. The conformation of the 4H and EF-hand domains is essentially identical in the two structures. In the absence of phosphopeptide, the SH2 domain makes little contact with the 4H domain and its position is likely to vary, as we observe slightly different conformations among the three molecules in the asymmetric unit. Phosphopeptide binding induces a domain 'closure', in which the SH2 domain rotates to pack against the helical domain, completing the phosphotyrosine-binding pocket, as in **d**. **d**, Molecular surface representation of the Cbl-N domain, coloured by domain. The 4H domain (yellow) forms a portion of the phosphotyrosine-binding pocket. Residues 289–297 of the bound ZAP-70 phosphopeptide are shown as a stick model. The three N-terminal residues in the peptide are disordered and are not included. In the liganded structure, about 1,200 Å<sup>2</sup> of the SH2 domain is buried as a result of interaction with the other two domains; 500 Å<sup>2</sup> is buried in the interface with the 4H domain, and 700 Å<sup>2</sup> is buried in the interface with the EF-hand. The 4H and EF-hand domains share a solvent-excluding interface of ~800 Å<sup>2</sup>.



**Figure 4** Mutations in the phosphotyrosine-binding pocket, the calcium-binding site and the 4H domain disrupt recognition of ZAP-70 by Cbl-N. Jurkat T cells were left untreated (–) or stimulated by anti-TCR crosslinking using OKT3 antibody (+). Lysates from resting or activated cells were probed with wild-type or mutant GST–Cbl-N bound to glutathione beads and the associated proteins were resolved by SDS–PAGE and immunoblotted for the presence of tyrosine-phosphorylated proteins using the anti-phosphotyrosine antibody RC20H (**a**). The same blot was stripped and re-probed with an anti-ZAP-70 antibody (**b**). Mutation of Gly 306 to Glu (corresponding to the loss-of-function mutation in Sli-1) or Arg 294 to Lys within the SH2 domain disrupt association of Cbl-N with ZAP-70. Mutations in the calcium-binding E2F2 loop, including Asp 229 to Gln and Glu 240 to Ser diminish recognition of ZAP-70 by Cbl-N. Substitutions for Ser 80 and Pro 82 in the AB loop of the 4H domain also prevent ZAP-70 precipitation.

the AB loop in the 4H domain, thereby completing the phosphotyrosine-binding pocket (Fig. 3d). A similar mechanism is employed in the tandem SH2 domains of ZAP-70, where residues of the abutting C-terminal SH2 domain complete the phosphotyrosine-binding pocket of the incomplete N-terminal domain upon binding appropriately spaced phosphotyrosine motifs<sup>22</sup>.

The primary specificity-determining interactions in the Cbl-N / ZAP-70 complex appear to be C-terminal to the phosphotyrosine. In particular, Pro 296 in the ZAP-70 peptide (pY+4) packs in a shallow hydrophobic cleft (Fig. 3b). This pocket might accept other medium-sized hydrophobic residues as well as proline. A glutamic acid residue at position pY+3 in the peptide also makes specific contacts. In contrast, residues N-terminal to the phosphotyrosine

are poorly ordered and we observe no electron density for the first three residues of the ZAP-70 peptide. A phosphopeptide library screen with Cbl-N indicated that the binding motif for the domain could be D(N/D)XpY<sup>6</sup>. The library varied in the three positions preceding and following the phosphotyrosine, and contained lysine residues at positions C-terminal to pY+3. The Cbl-N structure indicates that a hydrophobic residue at pY+4 is probably needed for high-affinity binding, so predictions made on the basis of the library screen may not be accurate. However, a preference for asparagine at pY-2, as predicted from the screen, is consistent with the observed interactions in the structure. Asp 290 (at pY-2) is poorly ordered, but makes an unexpected hydrogen bond to a phosphate oxygen. We would expect an asparagine residue at this position to make a more favourable interaction. Phosphorylation sites with aspartate or asparagine residues at pY-2 and hydrophobic residues at pY+4 are found in other Cbl-binding partners, including Syk and the EGF receptor.

Several features of the Cbl-N/ZAP-70 complex suggest that the 4H, EF-hand and SH2 domains together form an integrated structure that is crucial for phosphoprotein recognition. As we have discussed, the AB loop of the 4H domain forms a portion of the phosphotyrosine-binding pocket. The EF-hand domain roughly positions the 4H domain with respect to the SH2 domain in a way that seems to require calcium; the calcium-binding site in the EF hand is at the centre of an ~800 Å<sup>2</sup> interface with the 4H domain. Calcium coordination may define the orientation of the E2 and F2 helices, which form the interface with the SH2 domain. In calcium-regulated EF-hand proteins, binding of calcium to the EF loop stabilizes an 'open' conformation of the E and F helices that is required for interaction with target helical peptides<sup>7</sup>. In our structures, the calcium-bound E2F2 motif adopts an open conformation, and helix αN, which connects the EF-hand and SH2 domains in the primary structure, is positioned between the ends of the E2 and F2 helices in a position similar to that occupied by bound helical peptides in calmodulin structures.

To test whether the three domains form an integrated recognition module, we mutated key residues in each domain and compared the ability of wild-type and mutant proteins to precipitate ZAP-70 from activated Jurkat T-cell lysates. In the SH2 domain, substitution of lysine for the universally conserved 'FLVRES' arginine (Arg 294) disrupts the interaction of Cbl-N with ZAP-70, as does mutation of Gly 306 to glutamic acid (Fig. 4). This substitution corresponds to a loss-of-function mutation discovered in Sli-1<sup>13</sup> and disrupts the phosphotyrosine-binding function of Cbl and the transforming ability of v-cbl<sup>3,5</sup>. The mutation is in strand βC of the SH2 domain, near the phosphotyrosine pocket. The structure suggests that a

**Table 1** Data collection, phasing and refinement statistics

|                                                          | Cbl-N                                                         | Cbl-N MeHg derivative                                         | Cbl-N / ZAP70-292 complex        |
|----------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------|----------------------------------|
| Resolution (Å)                                           | 15–2.20                                                       | 15–2.24                                                       | 20–2.10                          |
| Space group                                              | C2                                                            | C2                                                            | P6                               |
| Unit cell (Å)                                            | <i>a</i> =159.96 <i>b</i> =105.48 <i>c</i> =84.92<br>β=92.06° | <i>a</i> =160.04 <i>b</i> =106.73 <i>c</i> =84.84<br>β=92.30° | <i>a</i> = 122.3 <i>c</i> =55.65 |
| Molecules/a.s.u.                                         | 3                                                             | 3                                                             | 1                                |
| <i>R</i> <sub>sym</sub> (%)                              | 6.7                                                           | 5.5                                                           | 7.9                              |
| Reflections (total/unique)                               | 178,845/70,918                                                | 186,473/67,069                                                | 35,952/20,620                    |
| Completeness (%)                                         | 99.3                                                          | 98.1                                                          | 76.9 (91% to 2.5 Å)              |
| <i>R</i> <sub>iso</sub> / <i>R</i> <sub>anom</sub> (%)   |                                                               | 35.1 / 6.5                                                    |                                  |
| Phasing power (centric/acentric)                         |                                                               | 1.11 / 1.51                                                   |                                  |
| <i>R</i> <sub>cryst</sub> (centric/acentric/anomalous %) |                                                               | 0.68 / 0.74 / 0.90                                            |                                  |
| FOM                                                      |                                                               | 0.41                                                          |                                  |
| Number of sites                                          |                                                               | 15                                                            |                                  |
| Refinement statistics                                    |                                                               |                                                               |                                  |
| Resolution range (Å)                                     | 15–2.20                                                       |                                                               | 20–2.10                          |
| Non-hydrogen atoms                                       | 8,092                                                         |                                                               | 2,919                            |
| Water molecules                                          | 669                                                           |                                                               | 358                              |
| <i>R</i> <sub>cryst</sub> / <i>R</i> <sub>free</sub> (%) | 22.1 / 28.7                                                   |                                                               | 17.3 / 24.6                      |
| R.m.s.d. bond lengths / angles                           | 0.018 Å / 2.1°                                                |                                                               | 0.013 Å / 1.84°                  |

glutamic acid at this position could form a buried salt bridge with Arg 294, preventing its interaction with phosphotyrosine. We disrupted the calcium-binding site in the EF hand by mutating acidic residues that participate in calcium coordination (Fig. 4). Mutation of Glu 240 to serine prevented the fusion protein of glutathione-S-transferase with Cbl-N (GST-Cbl-N) from interacting with ZAP-70, and mutation of Asp 229 to glutamine markedly weakened this interaction. In the 4H domain, we substituted aspartate for Ser 80 and alanine for Pro 82 to test the importance of the contribution of the AB loop to the phosphotyrosine-binding pocket. We detected no precipitation of phosphorylated ZAP-70 for either substitution. On the basis of the structure and our mutagenesis results, we conclude that the three domains together form the functional phosphoprotein-binding unit, and that calcium coordination plays a key structural role in phosphoprotein recognition.

Whether Cbl function is regulated *in vivo* by changes in intracellular  $\text{Ca}^{2+}$  concentration remains to be investigated. The affinity of Cbl for calcium may be sufficiently high that it is constitutively calcium-bound; no calcium was added to the unbound Cbl-N and the crystallization buffer included citrate, a weak  $\text{Ca}^{2+}$ -chelator. Addition of 1 mM EGTA to activated Jurkat lysates does not prevent precipitation of ZAP-70 by GST-Cbl-N (data not shown), indicating that calcium probably binds with high affinity. Thus, our *in vitro* results obtained with the isolated Cbl-N domain suggest that the calcium site is unlikely to serve a regulatory role. However, its regulatory potential depends upon its affinity for calcium in the cellular milieu; interactions with other domains in Cbl or with other proteins may alter its affinity for calcium. The architectural complexity of Cbl-N, in comparison with typical SH2 or PTB domains, suggests that it has a capacity for an additional binding or regulatory function. □

## Methods

**Expression, purification and crystallization.** Cbl-N was produced as a GST-fusion protein in *E. coli* strain DH5 $\alpha$  using the expression plasmid pGEX-4T-3 (Pharmacia). Cleared cell lysates were incubated overnight with glutathione-agarose beads. After extensive washing with PBS, the desired Cbl-N fragment was digested from the beads with thrombin (using a molar ratio of about 1:500) in 50 mM Tris, pH 7.4, 200 mM NaCl, and 2.5 mM calcium chloride. The digested Cbl-N was recovered and further purified by affinity chromatography on phosphotyrosine-Sepharose<sup>23</sup>. The purified protein was dialysed exhaustively against storage buffer (20 mM HEPES, pH 7.0, 200 mM NaCl, 2 mM DTT) and concentrated to 5 mg ml<sup>-1</sup> for crystallization in a Centricon (Amicon). The purified protein includes residues 25–351 of human Cbl, plus two residues (Gly-Ser) from the GST fusion.

Cbl-N crystals were obtained by combining 25  $\mu$ l protein and 25  $\mu$ l precipitant solution (20% PEG 4000, 100 mM sodium citrate, pH 5.6, 200 mM ammonium acetate) in a sealed glass depression plate at 22 °C. The crystals were maintained in a cryo-stabilization buffer (22% PEG 4000, 100 mM sodium citrate, pH 5.6, 200 mM ammonium acetate, 200 mM NaCl, 20% glycerol) for at least 2 h before flash-freezing by plunging into liquid nitrogen. For co-crystallization, a 2-fold molar excess of the ZAP-70 pTyr-292 peptide TLNSDGPYTPEPA was added to Cbl-N at 5 mg ml<sup>-1</sup> in storage buffer. Crystals were grown in hanging drops at 22 °C by mixing 2  $\mu$ l protein/peptide solution with 2  $\mu$ l well solution containing 20% PEG 8000, 0.2 M calcium acetate, 2 mM DTT, and 0.1 M sodium cacodylate, pH 6.1. Cbl/ZAP-70 co-crystals were briefly dunked in a stabilizing buffer containing the well solution plus 20% glycerol, and flash-frozen in liquid nitrogen. All diffraction data were obtained by using a Quantum-4 CCD detector (ADSC) on the F1 beam line at CHESS (Table 1) at -165 °C. Diffraction images were indexed, integrated and scaled using DENZO and SCALEPACK<sup>24</sup> (Cbl/ZAP-70 complex) or MOSFLM<sup>25</sup> (unliganded structure).

**Cbl-N structure determination.** The unliganded Cbl-N structure was determined by SIRAS using a methyl mercury nitrate derivative (1 mM, overnight soak). Fifteen mercury sites were located by using difference-Patterson and difference-Fourier methods using the CCP4<sup>25</sup> program suite and the programs PATSOL (L. Tong) and HEAVY (T. Terwilliger). Structure-factor

phases were calculated using MLPHARE<sup>25</sup>, and improved with solvent-flipping in DM<sup>25</sup>. An electron density map calculated using these phases at 2.9 Å resolution revealed clear main-chain density and substantial side-chain detail. With the aid of skeletonization using BONES<sup>26</sup>, it was possible to trace the path of the main chain for one of the three molecules in the asymmetric unit (molecule B). Density for the other two molecules was broken and hard to interpret. The skeleton for molecule B was used to generate a molecular envelope for NCS averaging. Initial rotation matrices describing the relative orientations of molecules A, B and C were calculated from the heavy-atom sites (mercury bound to five sites in each of the three molecules). NCS averaging with DM<sup>25</sup> was used to improve phases at 2.7 Å resolution, and then extend phases to 2.2 Å, the limit of the native data set. The resulting electron density map was readily interpretable (Fig. 2). A model including residues 47–351 of each of the three molecules was built using the graphics program O<sup>26</sup>. No electron density is seen for 24 residues at the amino terminus. The model was refined using simulated annealing and positional refinement in X-PLOR<sup>27</sup>, with tight NCS restraints. The model includes 669 water molecules, which were positioned by using the program ARP (V.Lamzin). An overall thermal B factor and tightly restrained individual B factors were refined, and a bulk-solvent model was incorporated. Crystallographic *R* values and stereochemical parameters are presented in Table 1.

The structure of the Cbl-N/ZAP70 phosphopeptide complex was determined by molecular replacement with the program AmoRe<sup>28</sup>. Use of the complete unliganded Cbl-N structure as a search model yielded clear rotation and translation peaks, and maps phased with the appropriately positioned model revealed strong electron density for the 4H and EF-hand domains, but no interpretable density for the SH2 domain. The model was therefore broken into two fragments (residues 47–265 and residues 266–351), and the rotation and translation searches were repeated with each fragment. The 4H/EF-hand fragment yielded a solution essentially identical to that from the intact model. The SH2 domain was then positioned using translation searches conducted in the context of the appropriately positioned 4H/EF-hand fragment. After rigid-body and positional refinement using X-PLOR<sup>27</sup>, electron density maps calculated with the combined model revealed clear density for all domains, and readily interpretable density for the bound ZAP-70 phosphopeptide. After construction of the peptide, the structure was refined with iterative cycles of manual refitting and simulated annealing and positional refinement in X-PLOR (Table 1). Restrained individual temperature factors were refined. The model includes residues 47–351 of c-Cbl, residues 289–297 of ZAP-70, and 358 water molecules.

**Illustrations.** Figure 1a was prepared with MOLSCRIPT<sup>29</sup>, Fig. 2 with program O<sup>26</sup>, and Figs 1e and 3 with GRASP<sup>30</sup>.

Received 7 December 1998; accepted 19 January 1999.

- Thien, C. B. & Langdon, W. Y. c-Cbl: a regulator of T cell receptor-mediated signalling. *Immunol. Cell Biol.* **76**, 473–482 (1998).
- Liu, Y. C. & Altman, A. Cbl: Complex formation and functional implications. *Cell Signal.* **10**, 377–385 (1998).
- Miyake, S. *et al.* The Cbl proto-oncogene product: From enigmatic oncogene to center stage of signal transduction. *Crit. Rev. Oncogen.* **8**, 189–218 (1998).
- Wang, R. L. & Samelson, L. E. Complex complexes: signaling at the TCR. *Immunity* **5**, 197–205 (1996).
- Thien, C. B. & Langdon, W. Y. EGF receptor binding and transformation by v-cbl is ablated by the introduction of a loss-of-function mutation from the *Caenorhabditis elegans* *sl-1* gene. *Oncogene* **14**, 2239–2249 (1997).
- Lupher, M. L. Jr, Songyang, Z., Shoelson, S. E., Cantley, L. C. & Band, H. The Cbl phosphotyrosine-binding domain selects a D(N/D)XpY motif and binds to the Tyr292 negative regulatory phosphorylation site of ZAP-70. *J. Biol. Chem.* **272**, 33140–33144 (1997).
- Ikuura, M. Calcium binding and conformational response in EF-hand proteins. *Trends Biochem. Sci.* **21**, 14–17 (1996).
- Kuriyan, J. & Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288 (1997).
- Saurin, A. J., Borden, K. L., Boddy, M. N. & Freemont, P. S. Does this have a familiar RING? *Trends Biochem. Sci.* **21**, 208–214 (1996).
- Blake, T. J., Shapiro, M., Morse, H. C. & Langdon, W. Y. The sequences of the human and mouse c-cbl proto-oncogenes show v-cbl was generated by a large truncation encompassing a proline-rich domain and a leucine zipper-like motif. *Oncogene* **6**, 653–657 (1991).
- Galisteo, M. L., Dikic, I., Batzer, A. G., Langdon, W. Y. & Schlessinger, J. Tyrosine phosphorylation of the c-cbl proto-oncogene protein product and association with epidermal growth factor (EGF) receptor upon EGF stimulation. *J. Biol. Chem.* **270**, 20242–20245 (1995).
- Deckert, M., Elly, C., Altman, A. & Liu, Y. C. Coordinated regulation of the tyrosine phosphorylation of Cbl by Fyn and Syk tyrosine kinases. *J. Biol. Chem.* **273**, 8867–8874 (1998).
- Yoon, C. H., Lee, J., Jongeward, G. D. & Sternberg, P. W. Similarity of *sl-1*, a regulator of vulval development in *C. elegans*, to the mammalian proto-oncogene c-cbl. *Science* **269**, 1102–1105 (1995).
- Meisner, H. *et al.* Interactions of *Drosophila* Cbl with epidermal growth factor receptors and role of Cbl in R7 photoreceptor cell development. *Mol. Cell Biol.* **17**, 2217–2225 (1997).



15. Ota, Y. & Samelson, L. E. The product of the proto-oncogene *c-cbl*: a negative regulator of the Syk tyrosine kinase. *Science* **276**, 418–420 (1997).
16. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
17. Kretsinger, R. H. EF-hands embrace. *Nature Struct. Biol.* **4**, 514–516 (1997).
18. Essen, L. O., Perisic, O., Cheung, R., Katan, M. & Williams, R. L. Crystal structure of a mammalian phosphoinositide-specific phospholipase C $\delta$ . *Nature* **380**, 595–602 (1996).
19. de Beer, T., Carter, R. E., Lobel-Rice, K. E., Sorkin, A. & Overduin, M. Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science* **281**, 1357–1360 (1998).
20. Becker, S., Groner, B. & Muller, C. W. Three-dimensional structure of the Stat3 $\beta$  homodimer bound to DNA. *Nature* **394**, 145–151 (1998).
21. Chen, X. *et al.* Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell* **93**, 827–839 (1998).
22. Hatada, M. H. *et al.* Molecular basis for the interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. *Nature* **377**, 32–38 (1995).
23. Eck, M. J., Shoelson, S. E. & Harrison, S. C. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56<sup>lck</sup>. *Nature* **362**, 87–91 (1993).
24. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
25. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
26. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Brunger, A. *X-PLOR Version 3.0: A System for Crystallography and NMR* (Yale University Press, New Haven, 1992).
28. Navaza, J. (ed.) *AMoRe: A New Package for Molecular Replacement* (SERC, Daresbury, UK, 1992).
29. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

**Acknowledgements.** We thank C. Dahl for synthesis and purification of the ZAP-70 phosphopeptide, the staff at MacCHESS for assistance with data collection, and S. Harrison and T. Roberts for comments on the manuscript. M.J.E. is a recipient of a Burroughs–Wellcome Career Award in the Biomedical Sciences. Diffraction data were recorded at the Cornell High Energy Synchrotron Source (CHESS), which is supported by grants from the NSF and NIH.

Correspondence and requests for materials should be addressed to M.J.E. (e-mail: eck@red.dfc.harvard.edu). Atomic coordinates have been deposited with the Protein Data Bank (Brookhaven National Laboratory) under accession numbers 2cbl and 1b47.

## erratum

# Protein translation and folding are coupled by an endoplasmic-reticulum-resident kinase

Heather P. Harding, Yuhong Zhang & David Ron

*Nature* **397**, 271–274 (1999)

As a result of a production error, the arrows in Fig. 3d were inappropriately aligned. The correct figure is reproduced below.

